

# nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE



## FEELING THE HEAT

Sex reversal drives switch from genetic to environmental sex determination in the bearded dragon **PAGES 43 & 79**

### CLIMATE CHANGE

#### CAN THE TREES KEEP UP?

*An uncertain future for forest carbon sinks*

**PAGE 20**

### POLICY

#### TRAVELLING HOPEFULLY

*Six routes to better transport planning*

**PAGE 29**

### SOLAR SYSTEM

#### PIT STOP FOR ROSETTA

*Sinkholes on comet 67P a sign of subsurface activity*

**PAGES 42 & 63**

**NATURE.COM/NATURE**

2 July 2015 £10

Vol. 523, No. 7558





# THIS WEEK

## EDITORIALS

**FORENSICS** Mobile camera converted to colour-analysis kit **p.6**

**WORLD VIEW** How one institute is tackling gender inequity **p.7**



**MANE EVENT** Lions head back to Rwanda after 15 years **p.10**

## Success in failure

*A failed crop trial of genetically modified wheat still provides crucial lessons for those battling to provide the planet's growing population with a sustainable food supply.*

It is rare for failures to be lauded in science. History, as it is often said, is written by the winner. The history of research is no different.

But failure in science is vital. Another cliché about history is equally applicable to scientific flops: people who are ignorant of them are doomed to repeat them. Which brings us to a green — and to some, an unpleasant — field in England.

In 2012, a team based at Rothamsted Research, an agricultural-science institute a short train ride north of London, planted wheat that they had genetically modified to emit a chemical used by aphids as a warning that they are under attack. The researchers wanted to see whether this would give the crops a way of repelling the damaging pests. They thought that the chemical might also attract insect parasites alerted to the promised presence of aphids.

Before they got the chance, the crops attracted a swarm of protesters. Opponents of genetic modification (GM) technology mounted an imaginative, if sometimes bizarre, campaign against the trial, complete with dubious scientific claims, loaves of bread adorned with cartoon cow heads, and videos promising to 'Take the Flour Back' complete with rock-music soundtrack. The research itself cost £732,000 (US\$1.2 million) over five years. Securing the site from those who threatened to tear it up cost nearly £1.8 million.

The idea behind what has, rather unfortunately, become known as 'whiffy wheat' showed promise in the laboratory. Yet in field trials the crop is an unquestionable failure. A paper published on 25 June in the journal *Scientific Reports* notes that the GM crops "showed no reduction in aphids or increase in parasitism" compared with controls (T. J. A. Bruce *et al. Sci. Rep.* <http://doi.org/5sr>; 2015).

This is disappointing on many levels. First, because of the effort — and money — that has gone into the concept. Second, because GM crops will surely have a major role in providing a future sustainable food supply. As Earth's population grows, so does its appetite. Work aimed at increasing crop yields, by both GM and non-GM methods, is among the most crucial research being conducted on the planet. So hostility towards GM research — one reason why it is rare for such crop trials to reach field-scale studies in Europe — is still among the most important societal issues for science to address.

Some opponents of GM crops have reacted with predictable claims: that the trial was a waste of money, that investment in GM science should therefore be cut off, and that this one set-back means the entire concept is flawed. Hardly.

As with most negative results in research, things can still be learnt from this trial. The team might yet modify the way their crop emits the alarm pheromone and may experiment in areas with higher densities of parasites.

The crop failed, but so did the protests. The research was done; a useful result was obtained. Ironically, had the protests succeeded and the trial been abandoned, the protesters would be unable to crow about

the crop's failure. GM research continues at Rothamsted, as it does around the world. Some of it will work and some will not.

Those who wish to make an argument against GM crops face major problems. The rise of new techniques such as CRISPR means that what is and is not a GM organism is an increasingly grey area, both scientifically and for regulators.

And these crops, with all the controversy that comes with them, are no longer the sole preserve of huge agri-businesses. The use of GM technology is increasingly being passed to the people who really need it — those in developing countries who are trying to improve the agriculture of their nations.

**"Considering all GM crops as a single case is increasingly problematic."**

Considering all GM crops as a single case is increasingly problematic. Consumer-friendly traits, such as apples that do not turn brown, now vie with nutritional enhancement for developing nations and drought resistance. Small academic groups around the world are producing locally tailored varieties alongside the engineered staples that major companies sell in huge quantities to farmers in the developed world. And the debate is no longer limited to crops — on page 13, we report on GM pigs that could soon make their way into the human food chain.

All who care about evidence-based policy-making should thank those who continue to struggle against both the difficulties of doing science and the added difficulties caused by people who would see science abandoned. We will all need the fruits — and the cereals — of their labours. ■

## Gene politics

*US lawmakers are asserting their place in the human genetic-modification debate.*

The revelation in April that scientists had edited the genome of a human embryo — an inevitable development to anyone paying attention to biotechnological advances — has sparked the biggest bioethical debate of the year and one that will last for decades. The overwhelming consensus is that such embryos should not be brought to term in clinical settings — at least not for now. The debate over when, if ever, that should take place has played out in the scientific literature in duelling articles, arguments about the technology's efficacy and calls for an Asilomar-like conference on bioethics.

So it is little surprise that lawmakers are weighing in. On 16 June, a subcommittee of the US House Committee on Space, Science and



Technology held a hearing on human gene editing with witnesses who included Jennifer Doudna, a biochemist at the University of California, Berkeley, who was one of the inventors of the genome-editing system CRISPR, and the Institute of Medicine (IOM) president Victor Dzau.

The climate was more educational than controversial, with lawmakers asking the usual questions about the risks, benefits and ethics of engineering future generations of the human race. Parallels were drawn with another ongoing debate over 'three-parent embryos', in which an egg cell's diseased mitochondria are replaced with healthy mitochondria from a second woman. A decision on whether to allow that procedure in the United States is in the hands of the US Food and Drug Administration (FDA), which has commissioned an IOM report on the topic that is due this winter.

While the research and technology subcommittee grilled the experts, a separate subcommittee — of the House Appropriations Committee that funds the FDA — was meeting elsewhere on Capitol Hill to draft the agency's 2016 budget. The subcommittee wants to take no chances with human modification: a bill that it released on 17 June bans the FDA from using public funds to evaluate applications for clinical trials involving genetically modified human embryos. Ironically, the current wording could backfire: applications for permission to investigate new drugs are automatically approved in 30 days unless the FDA blocks them, which would require funds.

If the budget passes, this clause would be the first time that lawmakers have used the FDA to limit human embryo research. A 1996 law known as the Dickey–Wicker Amendment bans the use of federal funds to create human embryos for research, but does not pertain to FDA regulation. The National Institutes of Health (NIH) reaffirmed in April that heritable genetic modification falls under the Dickey–Wicker rule, and director Francis Collins said that clinical application of such technology is “viewed almost universally as a line that should not be crossed”.

Nevertheless, Congress is determined to have a say. Deeply embedded in a report accompanying the appropriations bill are orders from the funding committee that the FDA appoints “an independent panel of experts, including those from faith-based institutions with expertise on bioethics and faith-based medical associations” to evaluate

the IOM three-parent embryo report when it is released and to report back.

Although the FDA budget is far from becoming law — after undergoing another round of editing, it must still be passed by Congress and the Senate and signed by the president — the implication is clear. The powerful spending committee that holds the purse strings wants to be involved in the debate: an understandable and indeed necessary position.

**“This clause would be the first time that lawmakers have used the FDA to limit human embryo research.”**

Still, even allowing for political posturing, the apparent pre-emptive distrust of the IOM's highly respected peer-review process is alarming. The perennially underfunded FDA has already spent US\$1.17 million on the IOM committee, and although no budget is set aside for the new panel, it will soak up money that could be spent elsewhere. Also worrisome is the religious language, which harks back to 2010 when a court ruled in favour of religious

organizations' interpretation of Dickey–Wicker and briefly shut down all NIH-funded human embryonic stem-cell research.

As this journal has said, all voices, including those of faith-based groups, should be heard in the debate over human-genome editing; indeed, the input of highly influential religious groups is essential to make a decision on how and if to regulate, especially in the United States. But the IOM committee already includes a professor of religious studies — so why duplicate the effort? This mandate to the FDA is not one that should come from a secular government, which seems to be seeking to impress conservative supporters. As one ethicist put it: “It is a sign that the culture wars aren't dead.”

When it comes to human-genome editing, however, those wars are a reality that all must face — and that is a good thing. This opening salvo from Congress shows just how complex the coming debate over human genomic modification will be. Academics have spent the past months debating among themselves how to proceed with research and clinical applications, sometimes acting as though they will be the arbiters of the final decision. As public awareness of the technology increases, that ethical discussion will rightly be taken out of their hands alone and planted firmly in those of broader society. ■

# Light detective

*Smartphone camera set to come to the aid of sleuths, scientists and wine lovers.*

As any reader of detective fiction knows, a crucial clue needed to solve a murder is the time of death. The hero detective, typically, is frustrated by the vague responses of the forensics team: “sometime between Tuesday night and Thursday morning” does little to narrow down the range of suspects. Scientists have long tried to help. And forensic science, with DNA analysis at the forefront, now ensures that more real-life criminals can expect a knock on the door, sometimes decades after they thought they had evaded detection. However, it remains effectively impossible to accurately judge the age of a bloodstain. Corpse excluded, bloodstains are typically the most common piece of evidence encountered at a homicidal-crime scene.

Colour could be the key. After blood leaves the body it starts to dry; as it does so, it changes from red to brown. Back in 1907, the Italian researcher Louis Tomellini produced a chart of 12 bloody spots, to illustrate this colour change over a year. As forensic science developed through the twentieth century, so did bloodstain analysis. By the 1960s, researchers were using photospectrometry, recording reflectance spectra and working out how the rate of Tomellini's colour changes could be affected by different atmospheric conditions. These are useful

observations, and forensic analysis of the colour of bloodstains is today a common part of the forensics team's work. But the results are still too variable for the analysis to stand up in court.

Colour provides more-useful data than many might think. Spectrometry is a valuable technique in many areas, from drug discovery to environmental monitoring. And astronomers use spectrometers to probe the atmospheres of distant exoplanets for conditions that might support life. Spectrometers, in other words, have become indispensable instruments. But they tend to be expensive, complex machines. The most precise can also be bulky, making them difficult to use in the field. On page 67 of this issue, scientists describe a possible step forward. They have built an optical spectrometer that is both small and powerful, and potentially cheap enough to find use in consumer electronics — to detect corked wine perhaps.

Like many modern images, those analysed by the scientists are taken with a smartphone camera. These are selfies from the quantum world: the camera is converted into a spectroscope using suspensions of particles called colloidal quantum dots. Exposed to light, these tiny particles produce vivid colours, with the shade and hue determined by the particle size. With the right mixture of particles, a coating can be applied that can filter and analyse the wavelengths (and so colour) of incoming light.

The research is discussed in a News & Views article on page 39, which describes how it could be used to produce “ubiquitous sensing elements in household devices connected to the Internet”. Beware would-be bloody criminals, your fridge is watching you. ■

➔ **NATURE.COM**  
To comment online,  
click on Editorials at:  
[go.nature.com/xhunq](http://go.nature.com/xhunq)



WALTER AND ELIZA HALL INST.



## Practical policies can combat gender inequality

*Mechanisms to help researchers to balance work and home lives have made a positive difference to the gender balance at my institute, says Douglas Hilton.*

**H**ow can science address the gender-inequality problem? It is a persistent issue that has been highlighted again by the controversy over the recent comments by Nobel laureate Tim Hunt about his “trouble with girls”?

The problem in biomedical research was starkly demonstrated to me just before I became director of the Walter and Eliza Hall Institute of Medical Research in Melbourne, Australia, in 2009. I chaired my first meeting of the senior academic staff and, despite having had a high-profile female director — Suzanne Cory — for more than a decade, none of the 20 department heads or professors in the room were women.

I pledged to improve the gender balance, and five years on, I think we have made some progress. We now have four female professors or department heads. That is hardly a reason for wild celebration, but given that we began from such a woeful base it is a start.

So what have we done? Simply, we asked the people affected — women in their postdoctoral period — for their ideas.

For our institute, some of the simplest changes included steps to ensure that all important meetings are held within school hours, to make sure that researchers with child-care duties can attend.

We have also set up a dedicated office with hot-desks and an adjoining room in which small children can play and older children can do homework or watch television, under the supervision of their parents.

And we designated a separate room to allow women to breastfeed their infants or to express milk. The idea of women expressing milk in a toilet or a sick room — as was done before — seems as inappropriate as having a researcher making their coffee there.

What else? We demand that at least half of speakers at all conferences and workshops organized by the institute are women. And we created a gender-equality committee, with men and women, to monitor implementation of policies, gather data on progress and challenge us with new ideas.

That was the easy stuff. Some steps required more thought, major investment and time. The trend over the past 30 years of postponing scientific independence by having researchers work for longer as postdocs is generally problematic, but especially difficult for many women, because those career-defining years overlap with child-bearing years. Female postdocs are placed in an invidious position: take some time off and have your productivity drop to near-zero for a period, or postpone having children in the hope of obtaining a faculty position.

So we deliberately started to appoint faculty members at a younger age, in their early to mid-thirties, perhaps after a 2–4 year period as

postdocs. This provides women with resources they can use (postdocs, research assistants and students of their own) should they take time out from full-time work to have children and to care for them.

For women who have children during their postdocs, we offer technical support, paid for by the institute to make sure that their projects progress while they are on maternity leave.

We introduced a 5-year, Aus\$1.25-million (US\$960,000) fellowship to support a female laboratory head, who can spend the money as she wishes. It can pay for salaries, for instance, or for consumable expenses. And, given that the high cost of child care can prevent women from returning to work, the institute helps to pay for it — up to Aus\$15,000 each year for female postdocs and lab heads with pre-school-age children. Yes, men pay for child care too, but we have a surfeit of male lab

heads, and we cannot afford to do it for everyone.

We also pay for our female scientists to take children and a carer with them to academic conferences, both here and abroad. This can cost hundreds or sometimes a few thousand dollars, but we think that presenting at meetings is important for career development. We also pay for a ‘family room’ at local conferences to allow researchers to listen to talks while accompanied by their children — which is good for both men and women.

We want to do more. We are planning an on-site child-care centre and new fellowships to support women returning after extended leave. And we are considering making the lab-head role more flexible. Could it be done as a job-share, for example, with two faculty members splitting supervisory responsibilities, each working three days a week?

We know that these steps have made a difference. Some are expensive, but the ‘my-institute-has-no-money’ argument is rarely a good excuse for inaction. Every institution has some discretionary money and can choose to spend it in these ways rather than, say, on over-generous recruitment packages for well-established (usually male) scientists.

Bigger changes have occurred when we have spoken openly, passionately and sometimes bluntly about our situation and the challenges faced by women more broadly in Australian academia. The Australian Academy of Science has become a leader in gender-equality discussions. The Australian Academy of Technological Sciences and Engineering has undergone a similar cultural change. And the National Health and Medical Research Council has issued guidelines and minimum standards on gender equality to institutions that wish to receive funding. This is progress. ■

**Douglas Hilton** is director of the Walter and Eliza Hall Institute of Medical Research in Melbourne, Australia.  
e-mail: [hilton@wehi.edu.au](mailto:hilton@wehi.edu.au)

**WE DEMAND THAT  
AT LEAST  
HALF OF  
SPEAKERS  
AT CONFERENCES  
ORGANIZED BY THE  
INSTITUTE  
ARE WOMEN.**

➔ **NATURE.COM**  
Discuss this article  
online at:  
[go.nature.com/dr3vom](http://go.nature.com/dr3vom)



# RESEARCH HIGHLIGHTS

Selections from the  
scientific literature

## ECOLOGY

### Roadkill yields panther numbers

By counting the number of endangered panthers hit and killed by cars in Florida, researchers have estimated the population size of this rare cat. They say that it is the first statistically robust population estimate for the animals across their breeding range.

Brett McClintock of the National Marine Mammal Laboratory in Seattle, Washington, and his team used data on reported collisions with *Puma concolor coryi* (fewer than 20 per year) and traffic densities, as well as information from a small number of radio-collared panthers to estimate the total population across the state. They show that panther numbers seem to be slowly increasing, but may never have exceeded 150 individuals between 2000 and 2012.

This method could be applicable to other rare animals, the researchers say. *J. Appl. Ecol.* <http://doi.org/5sg> (2015)

## MARINE BIOLOGY

### Corals inherit love for heat

Heat tolerance in corals can be passed down the generations, suggesting that corals can adapt as the climate warms.



Researchers have suggested that corals physiologically acclimatize to higher temperatures rather than inherit heat tolerance. To test this idea, Line Bay at the Australian Institute of Marine Science in Townsville, Australia, Mikhail Matz at the University of Texas at Austin and their team bred corals (*Acropora millepora*; pictured) from two locations in Australia separated by 5° of latitude. Offspring produced by parents from the warmer area had an up to 10 times greater chance of

survival when exposed to heat, compared with the larvae of parents from the cooler region. Larvae generated by crossing corals from the warm area with those from the cool region inherited key genetic differences associated with heat tolerance.

Corals that thrive in the heat could be moved to other latitudes so that they reproduce with local corals and introduce heat tolerant adaptations to the population. *Science* 348, 1460–1462 (2015)

## NEUROBIOLOGY

### Target neurons to relieve asthma

Silencing signals from pain-sensing nerve cells in the lungs reduces the symptoms of asthma in mice.

When stimulated by allergens, these neurons cause airways to constrict and trigger symptoms such as coughing and wheezing. Bruce Levy and Clifford Woolf of Harvard Medical School in Boston, Massachusetts,



## ATMOSPHERIC SCIENCE

### Air pollution triggers floods

A catastrophic 2013 flood in China was probably caused, in part, by air pollution.

In July 2013, heavy rainfall resulted in a devastating flood in the mountains northwest of the Sichuan Basin in China (pictured). The basin has seen increasing industrial activity in the past few decades. Jiwen Fan at the Pacific Northwest National Laboratory in Richland, Washington, and her team modelled the region's atmospheric processes during the storm using different levels of aerosol emissions. By setting the modelled emissions

at a level similar to that before China's economic boom, the team found that the rate of rainfall in the storm would have been up to 60% lower than under current emissions levels. Aerosols trapped in the basin warm the air and suppress convection, allowing excess moisture to build up and condense into rain as it rises up over the mountains.

The authors suggest that future severe floods in the region could be mitigated by reducing air pollution, particularly black carbon.

*Geophys. Res. Lett.* <http://doi.org/5q9> (2015)

IMAGINECHINA/CORBIS

M. MATZ



and their colleagues blocked the activity of these cells and found that this reduced airway inflammation by reducing the production of immune-signalling molecules such as IL-5. The team reports that IL-5 triggers pain-sensing neurons to release a peptide called VIP that stimulates immune cells, creating a feedback loop that sustains allergies.

The results reveal a potential way to treat asthma and respiratory allergies.

**Neuron** <http://doi.org/5rf> (2015)

## MATERIALS

## DNA glues particles together

Researchers have assembled micrometre-sized particles into a variety of crystals using DNA as 'glue'.

DNA has been used to control the assembly of DNA-coated nanoparticles, but doing this with larger particles leads to the formation of random clumps that do not crystallize. To solve this, Marcus Weck, David Pine and their colleagues at New York University attached many short DNA strands to the surface of polymer particles. The high density of DNA strands — 5 to 25 times higher than in previous work — along with their short 'sticky' ends and the smooth particle surface resulted in the particles self-assembling into various crystalline designs.

The method could be used to make more complex structures out of a range of materials including metals and semiconductors, the authors say.

**Nature Commun.** 6, 7253 (2015)

## ASTRONOMY

## Bounty of dark galaxies found

Astronomers have discovered more than 850 faint galaxies in a galaxy cluster that could be made mostly of dark matter.

Using archived images from the Subaru Telescope in Hawaii, a team led by Jin Koda at Stony Brook University

in New York searched for observations of the Coma galaxy cluster, which is roughly 101 million parsecs (330 million light years) away. The team found 854 ultra-diffuse galaxies, a class of faint galaxy that can be as large as the Milky Way, but which has only 0.1% the number of stars. For these galaxies to remain gravitationally bound together, the researchers show that more than 99% of their mass must be dark matter.

This suggests that the crowded environment sucks gas away from these galaxies, leaving them largely unable to form stars.

**Astrophys. J. Lett.** 807, L2 (2015)

## NEUROSCIENCE

## Male mice process pain differently

Male and female mice use different types of immune cell to process chronic pain.

Studies of male mice have shown that immune cells called microglia in the spinal cord have an important role in chronic pain. To see whether this is the same in female mice, a team led by Jeffrey Mogil at McGill University in Montreal and Michael Salter at the University of Toronto, both in Canada, induced chronic pain in both sexes. The team then used drugs or antibodies to reduce microglia function. Whereas pain responses were reduced in the males, females were unaffected and instead recruited a different type of immune cell, called a T cell. This difference was linked to testosterone, which could make T cells less able to mediate pain in the males, leading to their use of microglia instead.

**Nature Neurosci.** <http://dx.doi.org/10.1038/nn.4053> (2015)

## ASTRONOMY

## 'Tatooines' may be common

Planets orbiting a binary star system — like Tatooine, the fictional home planet of Luke Skywalker in

## SOCIAL SELECTION

Popular topics  
on social media

## A call to fund people not proposals

Laboratory heads today spend too much time struggling to win funding from the US National Institutes of Health (NIH), and this pressure to fund raise is driving young scientists away, according to a much-discussed commentary in *Cell*. To address this problem, Ronald Germain, chief of the laboratory of systems biology at the National Institute of Allergy and Infectious Diseases (NIAID) in Bethesda, Maryland, argues that the NIH should make funding decisions based almost entirely on researchers' past accomplishments, and not on their future plans for specific projects. Irakli Loladze, a quantitative ecologist at the University of Maryland University College in Adelphi, tweeted "person-not-project"-based scheme can be game changer in how science is funded." Sally Rockey, director of the NIH Office of Extramural

Research, says that the agency is already taking steps to streamline the funding process and to support scientists despite an ever-tightening budget.

**Cell** 161, 1485–1491 (2015)

### ➔ NATURE.COM

For more on  
popular papers:  
[go.nature.com/kwvdu9](http://go.nature.com/kwvdu9)

*Star Wars* — could form with surprising ease.

Most known circumbinary planets orbit close to their stars, where the competing gravitational forces from the two stars make the orbits of nearby objects unstable or intersect. This prevents debris from clumping together to form planets. But Benjamin Bromley of the University of Utah in Salt Lake City and Scott Kenyon of the Smithsonian Astrophysical Observatory in Cambridge, Massachusetts, show with simulations that a zone exists near the host stars where the orbits of debris wobble, but do not cross, allowing for planet formation.

This suggests that Earth-sized 'Tatooines' could be common and that more are likely to be discovered soon.

**Astrophys. J.** 806, 98 (2015)



the world's appetite for seafood.

Reg Watson at the University of Tasmania in Taroona, Australia, and his colleagues analysed global fisheries, and seafood import and export data. They found that the minimum distance between where seafood is sourced and where it is consumed increased nearly sixfold from 1950 to 2011. Humans are now exploiting nearly 40% of the ocean's primary productivity, up from roughly 15% in 1950. The team predicts that the world's growing demand for seafood will be met only until about 2050, unless changes are made in marine farming.

**Nature Commun.** 6, 7365 (2015)

## FISHERIES

## Farming footprint is rapidly growing

Humans are venturing farther across the oceans and harvesting a greater proportion of the ocean's biomass to feed

### ➔ NATURE.COM

For the latest research published by  
Nature visit:

[www.nature.com/latestresearch](http://www.nature.com/latestresearch)

# SEVEN DAYS

The news in brief

## POLICY

### EPA overruled

On 29 June, the US Supreme Court struck down a regulation to limit mercury emissions from power plants. The court ruled 5–4 that the US Environmental Protection Agency (EPA) should have conducted a cost–benefit analysis when it decided to regulate mercury, instead of doing so later in the regulatory process. The EPA is likely to revise its regulation in accordance with the ruling. Most utility companies have already complied with the rule by installing emissions-reduction equipment or closing down old coal-fired plants.

### US–China talks

The United States and China agreed to continue to cooperate on several measures, such as cracking down on illegal trafficking of wildlife and nuclear materials, and extending their five-year bilateral clean-energy research project. The seventh Strategic and Economic Dialogue, held on 23–24 June in Washington DC, included wide-ranging diplomatic discussions about climate, security and trade.

### Research conflict

The Smithsonian Institution in Washington DC said on 26 June that it plans to adopt stricter procedures governing conflicts of interest for its academics. The institution reviewed its policies in February, after allegations that solar physicist and climate-change sceptic Willie Soon of the Harvard–Smithsonian Center for Astrophysics in Cambridge, Massachusetts, failed to disclose his industry funding on his publications. The Smithsonian says that it will review sponsored awards for conflicts of interest and adopt standard terms and



BRENDON CREMER/NS/MINDEN/CORBIS

## Lions restored to Rwanda after 15 years

Seven lions from South Africa began their journey to Akagera national park in Rwanda last week. The big cats — donated by two parks in KwaZulu-Natal province, where there is a surplus of lions — were selected for their reproductive

potential to repopulate Akagera. The last lion died in the Rwandan park 15 years ago, after the population was poisoned by cattle herders in the wake of the country's 1994 genocide when the park was left unmanaged.

conditions for this. It will also automate its annual financial-disclosure programme for all researchers and adopt a policy that requires scientists to disclose all sources of funding when publishing.

### Alberta carbon tax

The provincial government of Alberta, Canada, is to gradually double its carbon tax from Can\$15 (US\$12) per tonne to Can\$30 per tonne by 2017. Environment minister Shannon Phillips announced the plan on 25 June. She also appointed Andrew Leach, an environmental economist at the University of Alberta in Edmonton, to lead a panel that will review Alberta's overall climate policy before the United Nations climate summit in Paris this December.

## FUNDING

### Cancer campaign

Oregon Health & Science University in Portland announced on 25 June that it has raised US\$1 billion for its cancer centre in under 2 years. Philip Knight, co-founder of the sportswear company Nike, issued a challenge to the university in 2013: Knight would donate \$500 million only if the centre raised the same amount in 2 years. The money is thought to be the largest successful 'challenge pledge' in the United States and will go to research into cancer detection. See page 14 for more.

### Controversial fund

The European Council agreed on 25 June to create a European Fund for Strategic Investments

(EFISI), which will support large projects — including research — that face financing bottlenecks. The EFISI concept alarmed European researchers when it was first announced by the European Commission in January because some of its budget will be taken from the commission's Horizon 2020 research programme. The commission agreed in May to exclude some Horizon 2020 programmes, such as the European Research Council and the Marie Skłodowska-Curie postdoc programme, from contributing.

## EVENTS

### Climate group wins

In a surprise ruling, a Dutch court has ordered the Netherlands to take measures



RED HUBER/ORLANDO SENTINEL/TNS/GETTY

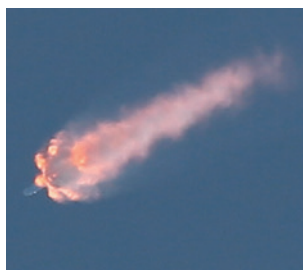
to cut its greenhouse-gas emissions by at least 25% by 2020, relative to 1990 levels. The court cited the undisputed state of climate science as evidence for its judgment, announced on 24 June. Under European Union pledges, the Netherlands needs to reduce its emissions by only 15%. But the Urgenda Foundation, a Dutch citizens' climate-change platform, said that this target is too low and sued the government for failing to take adequate action to prevent citizens from possible harm. The Dutch government can appeal against the ruling. See page 18 for more.

## Water worries

The water level of Lake Mead, the United States' largest reservoir, dipped to a historic low of 328 metres on 23 June, according to the US Bureau of Reclamation. That is lower than at any time since 1937. The site, located in Nevada, is part of the Colorado River system, which supplies water to roughly 40 million people in seven US states.

## SpaceX explosion

A SpaceX Falcon 9 rocket carrying supplies to the International Space Station (ISS) broke up (**pictured**) shortly after lift-off on 28 June from Cape Canaveral, Florida. The mission is the third supply run to the ISS to fail since



late October. "Preliminary analysis suggests the vehicle experienced an overpressure event in the upper stage liquid oxygen tank approximately 139 seconds into flight," said SpaceX, a private company in Hawthorne, California. The next supply mission to the ISS is scheduled for 3 July, when a Russian Progress spacecraft will launch from Kazakhstan. See [go.nature.com/wmrcm5](http://go.nature.com/wmrcm5) for more.

## DNA law tested

A US company is the first to face penalties under the Genetic Information Nondiscrimination Act, a US law that protects genetic privacy. Last week, a federal court in Georgia awarded US\$2.25 million to Jack Lowe and Dennis Reynolds, whose employer, Atlas Logistics Group Retail Services in Atlanta, Georgia, tested their DNA in a bid to identify who had left faeces on its premises. Neither man was the 'devious defecator'. See [go.nature.com/wtsvn3](http://go.nature.com/wtsvn3) for more.

## PEOPLE

### UCL stands ground

The head of University College London (UCL) has resisted calls to reinstate Nobel-prizewinning biologist Tim Hunt, who resigned from an honorary professorship after a media furore over comments he made about working with women. Michael Arthur, provost of UCL, said in a statement on 26 June: "Sir Tim has apologised for his remarks, and in no way do they diminish his reputation as a scientist. However, they do contradict the basic values of UCL — even if meant to be taken lightly — and because of that I believe we were right to accept his resignation."

## BUSINESS

### Patent tussle

Drug giant Eli Lilly has won a key UK patent fight over its blockbuster cancer drug Alimta (pemetrexed), the company announced on 25 June. Alimta's patent expires at the end of 2015, but Lilly, based in Indianapolis, Indiana, holds additional patents on a regimen of vitamins that are given with the drug. Dublin-based Actavis (now known as Allergan) proposed marketing an alternative form of Alimta after 2015, but the UK Court of Appeal ruled that this would violate Lilly's vitamin patents.

## COMING UP

### 6–9 JULY

Research on space planes, scramjets and other hypersonic vehicles will be discussed at the American Institute of Aeronautics and Astronautics conference in Glasgow, UK. [go.nature.com/hsvavt](http://go.nature.com/hsvavt)

### 7–9 JULY

Commercial and academic researchers meet in Boston, Massachusetts, for the ISS R&D conference to discuss innovation on the International Space Station. Elon Musk of SpaceX is among the speakers. [go.nature.com/dxkff2](http://go.nature.com/dxkff2)

### 9–13 JULY

The US Society for Developmental Biology holds its annual meeting in Snowbird, Utah. [go.nature.com/iy5m29](http://go.nature.com/iy5m29)

The decision could allow Lilly to maintain exclusivity in Britain and some other European countries until 2021.

## RESEARCH

### LIGO milestone

The Laser Interferometer Gravitational-Wave Observatory (LIGO) has tripled the sensitivity of its detectors, members of the LIGO collaboration said at a meeting last week in Waterloo, Canada. The team estimates that the improvement, revealed last month by an 'engineering run' of the upgraded detectors, will give LIGO a one-in-three chance of detecting gravitational waves during an observation run this autumn. LIGO has detectors in Washington state and Louisiana.

### NATURE.COM

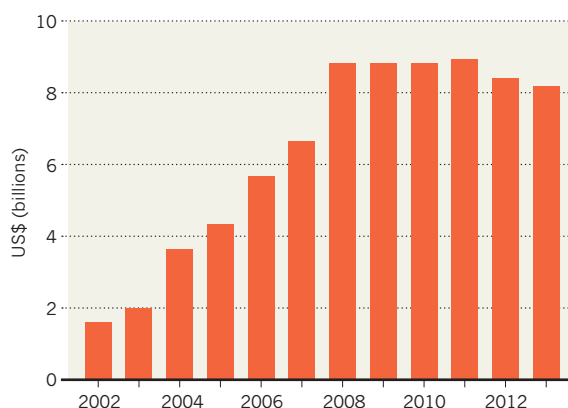
For daily news updates see: [www.nature.com/news](http://www.nature.com/news)

## TREND WATCH

Global efforts to combat AIDS must be scaled up urgently, or the epidemic is likely to rebound, warns a report by the Joint United Nations Programme on HIV/AIDS and *The Lancet* Commission. The document says that the next 5 years are crucial to eliminating AIDS as a major public-health threat by 2030 (P. Piot *et al. Lancet* <http://doi.org/5sm>; 2015). The rate of new infections is not falling fast enough, and donations to fight the epidemic have levelled off, the report notes.

### AIDS FUNDS PLATEAU

International monies from donor governments to low- and middle-income countries to tackle HIV/AIDS have levelled off.

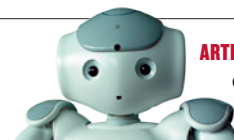
SOURCE: P. PIOT ET AL. LANCET [HTTP://DOI.ORG/5SM](http://doi.org/5sm) (2015)

# NEWS IN FOCUS

**FUND-RAISING** US cancer centre meets billion-dollar challenge **p.14**

**BOTANY** North America's herbaria an endangered species **p.16**

**CLIMATE** Dutch court orders extra emissions cuts **p.18**



**ARTIFICIAL INTELLIGENCE** The quest to build an ethical robot **p.24**

XI-JUN YIN



These meaty pigs could become the first genetically engineered animals to be approved for human consumption.

## GENE EDITING

# Super-muscly pigs created by small genetic tweak

*Researchers hope the genetically engineered animals will speed past regulators.*

BY DAVID CYRANOSKI

Belgian Blue cattle are hulking animals that provide unusually large amounts of prized, lean cuts of beef, the result of decades of selective breeding. Now, a team of scientists from South Korea and China says that it has created the porcine equivalent using a much faster method.

These 'double-muscled' pigs are made by disrupting, or editing, a single gene — a

change that is much less dramatic than those made in conventional genetic modification, in which genes from one species are transplanted into another. As a result, their creators hope that regulators will take a lenient stance towards the pigs — and that the breed could be among the first genetically engineered animals to be approved for human consumption.

Jin-Soo Kim, a molecular biologist at Seoul National University who is leading the work,

argues that his gene edits merely speed up a process that could, at least in principle, occur through a more natural route. "We could do this through breeding," he says, "but then it would take decades."

No genetically engineered animal has been approved for human consumption anywhere in the world, owing to fears of negative environmental and health effects. Fast-growing transgenic Atlantic salmon have languished in regulatory limbo for 20 years with the ►



► US Food and Drug Administration (see *Nature* **497**, 17–18; 2013).

Kim and his colleagues are part of a growing band of researchers who hope that gene editing, which can be used to disable — or knock out — a single gene, will avoid this. Reports of gene-editing applications in agriculture include the creation of hornless cattle. (Horns make the animals difficult to handle and are currently burned off in a painful procedure.) Researchers have also engineered pigs that are immune to African swine fever virus.

Key to creating the double-muscling pigs is a mutation in the myostatin gene (*MSTN*). *MSTN* inhibits the growth of muscle cells, keeping muscle size in check. But in some cattle, dogs and humans, *MSTN* is disrupted and the muscle cells proliferate, creating an abnormal bulk of muscle fibres.

To introduce this mutation in pigs, Kim used a gene-editing technology called a TALEN, which consists of a DNA-cutting enzyme attached to a DNA-binding protein. The protein guides the cutting enzyme to a specific gene inside cells, in this case in *MSTN*, which it then cuts. The cell's natural repair system stitches the DNA back together, but some base pairs are often deleted or added in the process, rendering the gene dysfunctional.

The team edited pig fetal cells. After selecting one edited cell in which TALEN had knocked out both copies of the *MSTN* gene, Kim's collaborator Xi-jun Yin, an animal-cloning researcher at Yanbian University in Yanji, China, transferred it to an egg cell, and created 32 cloned piglets.

Kim and his team have not yet published their results. However, photographs of the pigs “show the typical phenotype” of double-muscling animals, says Heiner Niemann, a pioneer in the use of gene-editing tools in pigs who is at the Friedrich Loeffler Institute in Neustadt, Germany. In particular, he notes, they have the pronounced rear muscles that are typical of such animals.

Yin says that preliminary investigations, show that the pigs provide many of the double-muscling cow's benefits — such as leaner meat and a higher yield of meat per animal. However, they also share some of its problems. Birthing difficulties result from the piglets' large size, for instance. And only 13 of the 32 lived to 8 months old. Of these, two are still alive, says Yin, and only one is considered healthy.

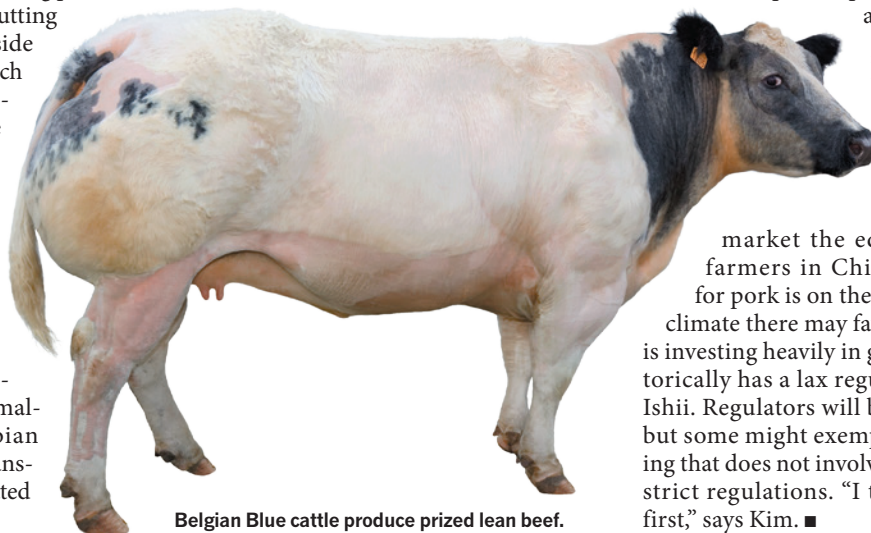
Rather than trying to create meat from such pigs, Kim and Yin plan to use them to supply sperm that would be sold to farmers for breeding with normal pigs. The resulting

offspring, with one disrupted *MSTN* gene and one normal one, would be healthier, albeit less muscly, they say; the team is now doing the same experiment with another, newer gene-editing technology called CRISPR/Cas9. Last September, researchers reported using a different method of gene editing to develop new breeds of double-muscling cows and double-muscling sheep (C. Proudfoot *et al. Transg. Res.* **24**, 147–153; 2015).

Because gene editing is a relatively new phenomenon, countries have only just started to consider how to regulate it in agricultural plants and animals. There are some signs that government agencies will view it more leniently than they do conventional forms of genetic modification: regulators in the United States and Germany have already declared that a few gene-edited crops fall outside of their purview because no new DNA has been incorporated into the genome. But Tetsuya Ishii, who studies international biotechnology regulation at the Hokkaido University in Sapporo, Japan, and who has done

an international comparison on GM regulations, says that gene editing will raise increasing alarm as it progresses in animals.

Kim hopes to market the edited pig sperm to farmers in China, where demand for pork is on the rise. The regulatory climate there may favour his plan. China is investing heavily in gene editing and historically has a lax regulatory system, says Ishii. Regulators will be cautious, he says, but some might exempt genetic engineering that does not involve gene transfer from strict regulations. “I think China will go first,” says Kim. ■



Belgian Blue cattle produce prized lean beef.

CLAUDIS THIRIET/PHOTONONSTOP/CORBIS

## FUNDING

# How an Oregon cancer institute raised a billion dollars

*Gains from two-year fund-raising frenzy will aid the early detection of tumours.*

BY HEIDI LEDFORD

Cancer researcher Brian Druker had no idea that a fund-raising gala would change his life. On 20 September 2013, armed with a speech that his wife had written for him, he waited patiently to be introduced by Philip Knight, the billionaire

co-founder of sportswear brand Nike.

Knight was a friend and benefactor; a few years earlier, he and his wife Penny had donated US\$100 million to the cancer centre that Druker directs at Oregon Health & Science University (OHSU) in Portland. But nothing had prepared Druker for what happened next. “Penny and I will donate

\$500 million to OHSU, if it is matched in pledges within two years in a fund-raising campaign,” Knight said, drawing gasps of surprise from the audience. “If the campaign raises \$499 million, we are relieved of our pledge,” he added. Druker turned in shock to his wife. “What do I do now?” he asked.

So began a frantic two-year scramble at the

OHSU Knight Cancer Institute to boost its fund-raising — about \$10 million in a good year — to \$250 million annually. On 25 June, OHSU announced that it had reached its target in 22 months. It is the largest amount a US institution has ever raised to win a challenge grant, according to the Indiana University Lilly School of Philanthropy in Indianapolis.

“Publicly we were always very confident, because if you aren’t, people aren’t going to donate,” Druker says. “But when we first got started, we thought, ‘How are we going to do this?’”

Billion-dollar campaigns are still relatively rare, says Bruce Flessner, a fund-raising consultant at Bentz Whaley Flessner in Minneapolis, Minnesota. And when universities do set out to raise that much, he notes, they typically take about seven years and dedicate the proceeds to all corners of the institution. The Knight Cancer Challenge aimed to fund a single institute at a university that is far from the clusters of wealth found in New York City or Silicon Valley.

“Portland is a great city, but it’s not minting billionaires at a fast rate,” says Flessner. “If there is a wealthy person in Oregon who hasn’t been asked to make a gift to that cancer programme, I would be shocked.”

## LOCAL APPEAL

But OHSU does have Druker, a renowned physician and researcher who made his name by laying the groundwork for the revolutionary leukaemia drug Gleevec (imatinib). The drug was approved by US regulators in 2001, and turned chronic myeloid leukaemia (CML) — once a death sentence for 70% of people diagnosed with it — into a long-term, manageable disease for 90% of patients.

Druker’s star power and Knight’s showmanship in designing and announcing the challenge galvanized the grass-roots fund-raisers. The campaign received more than 10,000 donations, given from 5 countries and every US state. Three-quarters of the money came from sources in Oregon. The largest single donation was \$100 million from Gert Boyle, chair of the Oregon-based company Columbia Sportswear. Boyle’s late sister, a molecular biologist, died of brain cancer and was a scientific mentor to Druker when he was an undergraduate at the University of California, San Diego.



Brian Druker spearheaded a campaign that raised US\$500 million in under 2 years.

The campaign decided early on to approach Oregon’s state legislature for \$200 million to construct two buildings for the cancer institute. OHSU pitched the expanded cancer centre to legislators as a way to create jobs for the state while fighting a disease that is the number-one killer of

***“If there is a wealthy person in Oregon who hasn’t been asked to make a gift, I would be shocked.”***

But two months later, the campaign hit a public-relations snag when its advertising — designed to be catchy and blunt — suggested that Druker’s work on Gleevec had “cured” CML. The pitches angered some people with CML, who must take expensive drugs for the rest of their lives while enduring side effects and the fear that their cancer will become resistant to treatment. Patients said that calling Gleevec a cure would slow the

Oregonians. When the state senate approved the measure by a vote of 28–2 in March 2014, Druker began to believe that Knight’s challenge could be met.

search for better therapies. Druker issued an apology and OHSU toned down the adverts to read: “That’s one cancer down. Now we’re going after other cancers as aggressively as they come after us.”

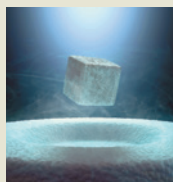
With the money now in hand, it is time for Druker, OHSU and the cancer centre to deliver on that promise. Druker aims to rapidly hire up to 30 principal investigators, and to provide researchers with a funding cushion intended to free them from the burden of constantly applying for grants. But the investigators will also be expected to meet research milestones. “We want to make progress as quickly as we can,” he says.

The institute will focus on detecting cancers early in their development, when treatments generally have a better chance of success. Druker also wants the institute to take advantage of emerging technologies to develop better tests that would reduce false diagnoses.

He is eager to turn his full attention to the science, but already feels nostalgic about the past two years. “It’s been busy,” he says. “But it’s been quite a ride.” ■



## TOP NEWS



Claimed superconductivity record is bolstered by magnetic data  
[go.nature.com/xihjqw](http://go.nature.com/xihjqw)

## MORE NEWS

- Researchers frustrated by failure to roll out ‘game-changing’ Ebola test [go.nature.com/7tvjwg](http://go.nature.com/7tvjwg)
- Conflict between the chromosomes that determine sex [go.nature.com/1xzdvp](http://go.nature.com/1xzdvp)
- Political science’s problem with research ethics [go.nature.com/tt4m4r](http://go.nature.com/tt4m4r)

## GRAPHENE



Andre Geim: Graphene’s buzz has spread  
[go.nature.com/sb4rov](http://go.nature.com/sb4rov)





Curator Vicki Funk of the US National Herbarium displays one of the collection's 5 million specimens.

## BOTANY

# Plant collections get pruned back

North America's herbaria wilt under budget pressure.

BY BOER DENG

Herbaria in North America are withering away. Collections of preserved plant specimens that have been accumulating for a century or more are being closed and consolidated as tight budgets and competition for space put pressure on universities to direct resources to facilities such as labs.

More than 100 North American herbaria have closed since 1997, leaving just over 600 remaining. The latest casualty came in May, when the University of Missouri in Columbia announced that it will close its Dunn-Palmer Herbarium, a 119-year-old collection of more than 170,000 plants and thousands of mosses, algae and fungi.

There is a perception that herbaria are dead places, says plant biologist Kathleen Pryer, who manages the herbarium at Duke University in Durham, North Carolina. But far from being relics, botanists argue, these repositories of preserved specimens are relevant to today's research.

For instance, DNA from specimen plants helps botanists to improve the accuracy of phylogenetic trees, and surveys of when and where specimens were collected can show the effects of climate change on species range. Ecologists and conservationists will always need to be able to distinguish thorn from thistle in the field, says biologist Roxanne Keller of the University of Nebraska Omaha. Digital archives are

useful, she says, but only with the real thing can you feel the points of a bristle or trace a tendril's curl. "You can't get those details from a picture."

That sensory experience may be less valued these days because many botanists now find themselves small players in broader biological-sciences departments. Few outside their field appreciate the merits of having specimens on hand. Department heads and deans are always "mystified" about herbaria, Pryer says.

Herbaria can feel more antique than avant garde. The US National Herbarium in Washington DC houses a preserved cutting from the first Concord grape, a commercially important US breed first cultivated in 1849. The label identifying a sunflower from South America, brown with age, is written in the spidery Cyrillic scrawl of a nineteenth-century Russian collector.

That fusty feel belies the present-day questions that the specimens are being tapped to tackle. Isotopic analyses of specimens of the rainforest species *Humiria balsamifera* that date as far back as 1788, for example, show that as atmospheric carbon dioxide levels increased with industrialization, the plants responded by increasing photosynthetic activity and using water more efficiently (D. Bonal *et al. Plant Cell Environ.* **34**, 1332–1344; 2011). The findings are important to climate modellers and others who want to predict how ecosystems will respond as CO<sub>2</sub> levels rise in coming decades.

Researchers have also used specimens collected during the first few decades of the

twentieth century to track the spread of cheat grass (*Bromus tectorum*), an invasive species from Europe, throughout the US southwest. The pattern supports a growing body of evidence that successful invasions require multiple introductions of an exotic species (A. R. Pawlak *et al. Biol. Invasion.* **17**, 287–306; 2015).

Herbaria do not necessarily disappear altogether when they close. Their specimens are often absorbed by other institutions: the Rancho Santa Ana Botanic Garden in Claremont, California, for example, has taken in at least three other collections since 2000. In one case, staff had to race against an impending rainstorm to rescue specimens from a loading dock where they had been unceremoniously dumped. "We more or less had to drop everything and go and fetch it," recalls Lucinda McDade, the garden's executive director.

Drama also surrounded the 2004 transfer of the herbarium at the University of Iowa in Iowa City to Iowa State University in Ames. A lawsuit tried to stop the move, but eventually more than 200,000 specimens were packed up and driven the 200 kilometers to Iowa State. And last year, the Brooklyn Botanic Garden in New York said that it would sell the building that housed its herbarium. In April, the garden lent its collection to the New York Botanical Garden until room can be found for it elsewhere. But some curators worry that the move out of Brooklyn will prove permanent.

Merging collections can have benefits, says James Miller, vice-president of science and conservation at the Missouri Botanical Garden in St Louis, which will absorb the Dunn-Palmer collection. Samples can be better curated at larger institutions, and might catch more researchers' attention. But taking in an orphaned collection is a mixed blessing. It can take years to catalogue the new samples, making it difficult to access them for study. And institutions must also find a way to do that work at a time of dwindling funds and staff cuts. "I'm glad we're getting new specimens," says Miller. "But a part of me is always sad when another herbarium closes."

Compared with other biological sciences, botanists feel that they have long struggled for respect. In 1988, 72% of the 50 top-funded US universities offered advanced degrees in botany. More than half of those programmes have been jettisoned, even though the need for soil and plant scientists is expected to rise modestly over the next decade, according to the US Bureau of Labor Statistics. "Getting people interested in living plants is a challenge," says Pryer. Convincing them of the importance of keeping flattened, wizened sprigs is even tougher.

But it is important to do so, botanists say. "A lot of us got started studying plants by wandering into the college herbarium by accident," says Vicki Funk, a botanist and curator at the US National Herbarium. "What happens if they all get carted off?" ■

CHRIS MADDALONI/NATURE

## MEDICAL RESEARCH

# Researchers pin down risks of low-dose radiation

Large study of nuclear workers shows that even tiny doses slightly boost risk of leukaemia.

BY ALISON ABBOTT

For decades, researchers have been trying to quantify the risks of very low doses of ionizing radiation — the kind that might be received from a medical scan, or from living within a few tens of kilometres of the damaged Fukushima nuclear reactors in Japan. So small are the effects on health — if they exist at all — that they seem barely possible to detect. A landmark international study has now provided the strongest support yet for the idea that long-term exposure to low-dose radiation increases the risk of leukaemia, although the rise is only minuscule (K. Leuraud *et al. Lancet Haematol.* <http://doi.org/5s4>; 2015).

The finding will not change existing guidelines on exposure limits for workers in the nuclear and medical industries, because those policies already assume that each additional exposure to low-dose radiation brings with it a slight increase in risk of cancer. But it scuppers the popular idea that there might be a threshold dose below which radiation is harmless — and provides scientists with some hard numbers to quantify the risks of everyday exposures.

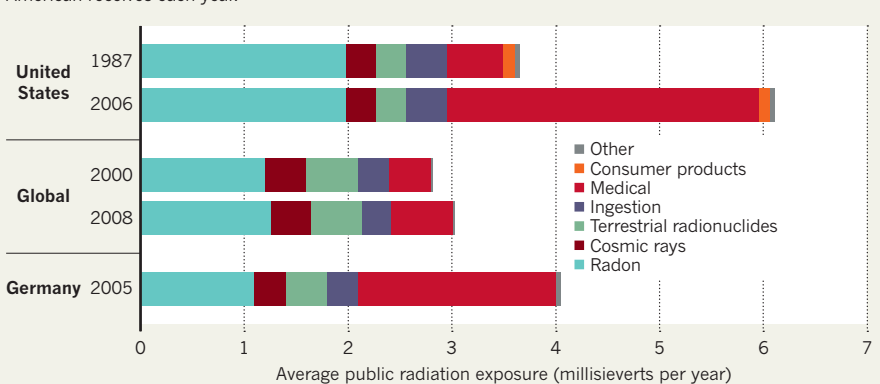
“The health risk of low-dose radiation is really very tiny, but the public is very concerned,” says Bill Morgan, who heads a systems-biology programme in low-dose radiation at the Pacific Northwest National Laboratory in Richland, Washington, and chairs the committee on radiation effects at the International Commission on Radiological Protection (ICRP) in Ottawa, Canada. That concern has driven a lot of investment in programmes trying to quantify the risk, he says. The European Commission, for example, has a 20-year road map to assess the problem. “We don’t do a very good job of explaining ourselves to the public, which finds it hard to put radiation risks in context — some people go to radon spas to treat their rheumatism while others won’t board planes for fear of cosmic rays,” he adds.

## RADIATION RISKS

Ionizing radiation — the kind that can pull electrons from atoms and molecules and break DNA bonds — has long been known to raise the risks of cancer; the higher the accumulated dose, the greater the damage. But it has proved extremely difficult to determine whether this relationship holds at low doses, because any

## RIISING BACKGROUND

A rise in medical scans over the past two decades has doubled the amount of radiation that the average American receives each year.



increase in risk is so small that to detect it requires studies of large numbers of people for whom the dose received is known. A study of more than 300,000 nuclear-industry workers in France, the United States and the United Kingdom, all of whom wore dosimeter badges, has provided exactly these data. A consortium of researchers coordinated by the International Agency for Research on Cancer (IARC) in Lyon, France, examined causes of death in the workers (one-fifth of whom had died by the time of the study) and correlated this with exposure records, some of which went back 60 years.

**“The health risk of low-dose radiation is really very tiny, but the public is very concerned.”**

The workers received on average just 1.1 millisieverts (mSv) per year above background radiation, which itself is about 2–3 mSv per year from sources such as cosmic rays and radon. The study confirmed that the risk of leukaemia does rise proportionately with higher doses, but also showed that this linear relationship is present at extremely low levels of radiation. (Other blood cancers also tended to rise with radiation doses, but the associations were not statistically significant.) The results were published on 21 June.

“It is a solid, unusually large study of individuals exposed to very low doses of ionizing radiation,” says epidemiologist Jørgen Olsen, director of the Danish Cancer Society Research Center

in Copenhagen. The finding implies that some cases of leukaemia will even be caused by a high level of natural background radiation, he adds, “though the increased risk for an individual is going to be vanishingly small”.

ICRP recommendations, which most national radiation-protection agencies follow, already call for monitoring of individuals whose annual exposure is likely to exceed 6 mSv. They restrict exposure to 20 mSv annually over 5 years, with a maximum of 50 mSv in any one year. Researchers expected that 134 of the workers (4.3 per 10,000 people) would die from leukaemia as a result of the average 27 years they spent in the industry; in fact, 531 people died from the disease. Even in this large study, there was no direct evidence that workers who had accumulated extremely low doses of radiation (below a total of 50 mSv) had an increased risk of leukaemia, says Olsen. But a mathematical extrapolation of the data suggests that each accumulation of 10 mSv of exposure raises a worker’s risk of leukaemia by 0.002%.

The data also challenge an ICRP assumption that accumulated low-dose exposure gives a lower risk of leukaemia than does a single exposure to the same total dose (based on the idea that the body has time to recover if the assault comes in tiny, spread-out doses). But such details are unlikely to change the overall ICRP recommendations, which are deliberately conservative, says Thomas Jung, from Germany’s Federal Office for Radiation Protection in Munich. ▶



## ► MEDICAL SCANS

A major, and increasing, source of low-dose radiation comes from the medical world, says David Richardson, an epidemiologist at the University of North Carolina and an author of the study. “The amount of radiation a US person receives in a year on average has doubled, mostly because of medical procedures,” he says (see ‘Rising background’). Computed-tomography (CT) scans are to blame for most of the rise; a typical abdominal scan delivers more than 10 mSv. Radiologist David Brenner of Columbia University in New York has calculated that of the 25 million people having CT scans in a year, 1 million will have accumulated more than 250 mSv over the previous 20 years.

One group that needs to pay particular attention to the findings are the tens of thousands of health workers who use radiological imaging to guide catheters through blood vessels of patients to reach into their hearts and brains, says Martha Linet, at the US National Cancer Institute’s radiation epidemiology programme in Bethesda, Maryland. These minimally invasive operative procedures are used ever more frequently, she says.

Epidemiological studies suggest that radiation exposure has health effects beyond cancer. The IARC-led consortium is now looking at the effect on solid cancers, and also on diseases such as heart attack and stroke. Other studies are under way to study the long-term impact of low-dose radiation on different cohorts. One, the Epi-CT study, is recruiting one million people from nine European countries who had CT scans as children; its analysis will be complete by 2017. In another, the Helmholtz Center Munich is analysing heart tissue from workers who died in the Mayak uranium mines in the South Urals, Russia.

Although the European Commission has been funding research on low-dose radiation for some time, equivalent programmes in the United States have stalled. In 2013, scientists wrote an open letter to the White House Office of Science and Technology Policy calling for renewed investment, and a bill is currently being debated in Congress calling for more work.

Getting funding for such studies is important, says Mike Atkinson, head of radiation biology at the Helmholtz Center Munich. Being able to quantify the effects of radiation will help doctors to balance risk against benefit when deciding whether to put children in CT scanners, he says. And further understanding the health impacts of low-dose radiation might aid decisions about how much remedial activity is needed to clean up soil contaminated by radioactivity from accidents or nuclear-power works, says Morgan. ■



In the Netherlands, concerns about rising sea levels have led citizens to sue to force emissions cuts.

## POLICY

# Courts weigh in on climate change

*Successful Dutch climate litigation may encourage action across Europe, but US courts seem unlikely to follow suit.*

BY QUIRIN SCHIERMEIER

A group of Dutch citizens weary of ineffectual climate diplomacy are celebrating after forcing change through legal action. Last week, following a lawsuit filed by a citizens’ climate-change platform called the Urgenda Foundation, a court in The Hague ordered the government of the Netherlands to cut greenhouse-gas emissions to at least 25% below 1990 levels by 2020 — substantially greater cuts than are required under the small country’s European Union (EU) obligations.

The ruling could encourage citizens of other countries to try using legal avenues to force stricter climate policies, says James Thornton, the London-based chief executive of Client Earth, an international group of environmental lawyers. “This is a very powerful decision

with possible far-reaching repercussions,” he says. “It is forcing the use of undisputed scientific results for responsible policy-making — a very remarkable step.”

The Dutch government may still appeal the ruling, and even if it does have to implement extra emissions cuts, these would barely dent global greenhouse-gas emissions. But the court made clear that although Dutch policy-makers can do little to reduce emissions in China or the United States, they still have an obligation to act out of a duty of care for their citizens.

Thornton hopes that other courts will judge similar lawsuits in the same way in future. One such case is pending in Belgium, which must reduce its emissions by only 15% below 2005 levels under current EU pledges. But it is unclear whether the landmark Dutch ruling, and any European lawsuits that might follow,

FRANS LEMMENS/ALAMY



will make waves in other parts of the world — particularly in the United States.

In 2007, the US Supreme Court authorized the Environmental Protection Agency (EPA) to regulate greenhouse-gas emissions that contribute to air pollution, because pollution could endanger public health or welfare. A series of greenhouse-gas reduction plans have followed. But attempts to get federal courts to order more-substantial cuts have so far come to nothing. Four years after the EPA decision, the Supreme Court rejected an effort by California and five other states to seek a cap on emissions from the utilities sector. The states argued that greenhouse gases are a ‘public nuisance’; however, the court countered that the EPA’s authority to regulate emissions prevented federal judges from using the public-nuisance argument. Attempts by others to claim liability against polluters and seek damages under civil law have also been unsuccessful.

### LIMITED POWER

In the United States, “there is no federal constitutional right to environmental protection”, says Richard Stewart, an environmental-law specialist at New York University. “Some state courts may recognize such a right, but the remedy might at best be limited to local sources.”

That seems to be the case in Washington state, where on 23 June, a Seattle court ordered

the state’s ecology department to reconsider a 2014 petition brought by eight school students to limit the state’s carbon dioxide emissions. The petition called for the agency to act in line with what scientific evidence says is needed to protect the climate and the environment. The agency initially denied the petition, but has been ordered to report back to the court by 8 July. Petitioners’ lawyer Andrea Rodgers, of the Western Environmental Law Center in

**“This is a very powerful decision with far-reaching repercussions.”**

Seattle, said that it was the first time a US court had ordered a state agency to consider the most current and best available climate science in deciding regulation on carbon emissions.

It would be unusual for a US court to demand a specific level of federal emissions regulation, as has happened in the Netherlands, says Michael Oppenheimer, who studies geosciences and international affairs at Princeton University in New Jersey. A court would be likely to do so only if there were a large gap between public safety and existing regulations, he says.

“If it became clear that US regulations, along with actions of other countries, are insufficient, then at some future date a court might invoke the objective to force stronger action,” he says. But, adds Oppenheimer, current US targets are

consistent with “at least some pathways” that would keep the world’s warming below 2 °C, the internationally recognized threshold for ‘dangerous’ climate change. ■

### CORRECTIONS

The News story ‘Election results delight scientists’ (*Nature* **522**, 264–265; 2015) stated that Gençay Gürsoy won a seat in the new Turkish parliament for the HDP. He did not; he is a member of the HDP assembly. The News story ‘Earth science wrestles with conflict-of-interest policies’ (*Nature* **522**, 403–404; 2015) erroneously stated that hydrologist Donald Siegel disclosed the provision of water samples by Chesapeake Energy Corporation only in a correction to his article. In fact, this information was included in the acknowledgements of his original paper.

### CLARIFICATION

The News story ‘Earth science wrestles with conflict-of-interest policies’ (*Nature* **522**, 403–404; 2015) did not make clear that Siegel’s findings related to gas production in general, and not just the process known as fracking. This has been clarified in the online version of the story.





# WEIGHING THE WORLD'S TREES

*Researchers are racing to determine whether forests will continue to act as a brake on climate change by soaking up more carbon.*

BY GABRIEL POPKIN

In a forest just west of Chesapeake Bay, Geoffrey Parker wraps a tape measure around a young tulip tree. He jots the reading down in a field notebook, marks the tree with blue chalk and moves on to the next trunk. Parker spends about 10 seconds on each tree. Wrap, measure, record. Since 1987, he and others have logged more than 300,000 tree measurements at their plots in the Smithsonian Environmental Research Center (SERC) near Edgewater, Maryland.

This 1,070-hectare site is filled with tulip trees, oaks, beeches and other mostly deciduous trees. Some stout specimens have stood here for centuries. Others are just a decade old, sprouting from land that was recently logged. To keep tabs on the growth, the researchers measure their trees every three to five years.

All that patient record-keeping can help to answer two major questions about climate change: how much carbon dioxide pollution are forests mopping up, and will their capacity shrink over time? Studies from Parker's group and others reveal that trees around the globe are going through a growth spurt and are absorbing billions of tonnes of the greenhouse gas, meaning that forests are putting a brake on global warming. But there is no guarantee that forests will keep that up, Parker says. "I think of it like these performance enhancers that some stellar athletes use: it bumps up performance, but not for ever."

In fact, studies of some regions suggest that forest growth may already be slowing down. And humans are adding to the problem by cutting down trees, especially in tropical forests. Getting an accurate reading on the status of Earth's forests is hard because scientists cannot wrap measuring tapes around the roughly 400 billion trees scattered across the planet. So researchers are exploring ways to track forest growth more efficiently, using planes and satellites. And they are feeding all of their data into sophisticated computer models that are designed to forecast how trees will respond in the future.

Such forest measurements are sorely needed as nations wrestle with how to slow climate change. Some plans call for wealthy governments or private companies to pay poorer nations in return for safeguarding the carbon in their forests. With a major

**Redwood trees can store carbon for more than 2,000 years.**

NATIONAL GEOGRAPHIC CREATIVE/GETTY IMAGES



international climate negotiation approaching later this year, and billions of dollars in forest payments potentially on the table, scientists are racing to advise countries and other stakeholders about just how much carbon trees are storing, and how long that carbon will stay locked up.

“The critical thing that matters is to what extent the biosphere remains a brake on the rate of global climate change,” says Yadvinder Malhi, a forest ecologist at the University of Oxford, UK. That brake will weaken or disappear if forests take up carbon more slowly. Worse, if forests start emitting more carbon than they absorb each year, they could become an accelerator. If that were to happen, says Malhi, “it makes it all the more challenging for us to bring CO<sub>2</sub> down to avoid some threshold of dangerous climate change”.

## THE MISSING SINK

In the 1990s, researchers stumbled across a mystery when they tried to track down all of the carbon humans were emitting by burning fossil fuels. Measurements showed that roughly three-quarters of the CO<sub>2</sub> was accumulating in the atmosphere and oceans. The remainder was presumably captured on land, but no one knew where it was going. The problem became known as the ‘missing sink’.

The world’s forests, which pull carbon out of the air through photosynthesis, were a possible hiding place. Today, they collectively hold around 650 billion tonnes of carbon, and it seemed plausible that they could be mopping up the missing carbon.

But ecologists were slow to acknowledge that forests could be the missing sink. The community’s reticence resulted largely from the work of pioneering ecologist Eugene Odum. He argued in the late 1960s that undisturbed ecosystems rapidly reach an equilibrium, after which they lose as much carbon through respiration, death and decay as they gain through photosynthesis<sup>1</sup>. Without much evidence to the contrary, Odum’s paradigm held sway for several decades. “Mathematically, it’s convenient if something is in equilibrium,” says Sebastiaan Luyssaert of the Laboratory for Climate Sciences and the Environment in Gif-sur-Yvette, France. “We were happy with it, because it made life easier.”

That started to change as ecologists analysed long-term data from big networks of forest research plots. Many of the measurements came from a trio of projects: the Amazon Forest Inventory Network (RAINFOR), the African Tropical Rainforest Observation Network (AfriTRON) and the Smithsonian’s Forest Global Earth Observatories (ForestGEO) network, which includes the SERC forest and 61 other plots around the world.

Starting in the late 1990s, scientists with the RAINFOR and AfriTRON networks began reporting that intact tropical forests were adding biomass, contradicting Odum’s hypothesis. At the Chesapeake site, Smithsonian ecologist Sean McMahon and his colleagues analysed 22 years’ worth of data and found that tree stands of all ages were growing two to four times faster than expected<sup>2</sup>. The tree growth records are backed up by CO<sub>2</sub> measurements taken on tall towers at more than 20 sites in North America and Europe: these ‘flux towers’ have revealed that many forests are absorbing more CO<sub>2</sub> than they are giving off<sup>3</sup>.

Researchers suspect several factors are at play. Because trees require CO<sub>2</sub> for photosynthesis, the atmospheric build-up of this gas can fertilize plants, allowing them to grow faster. Also, CO<sub>2</sub> warms the planet, which can lengthen the growing seasons of trees and speed up temperature-dependent processes involved in growth. Scientists are currently teasing out which factors have the largest roles.

Whatever the cause, all that accelerated growth is having a major effect on the global carbon cycle. In 2011, an international team led by US Forest Service researchers Yude Pan and Richard Birdsey concluded that the world’s trees had sequestered enough carbon during the period from 1990 to 2007 to account for the entire missing sink<sup>4</sup>. The hungriest carbon absorbers were the temperate forests, particularly areas where abandoned farmland had given way to young, fast-growing trees. High-latitude boreal forests ate up a smaller amount, and tropical forests, on balance, were not taking up carbon because tropical deforestation released about as much CO<sub>2</sub> as forests were soaking up. The team projected that if deforestation were halted, Earth’s forests could

take up around half of the carbon emitted by human activity, which would substantially slow down global warming.

But the uncertainties in these estimates are large because forest data are sparse and vary widely in quality. Many countries have no systematic forest inventory system or do not share their data. In their analysis, Pan and Birdsey relied largely on RAINFOR and AfriTRON for assessing the globe’s old-growth tropical forests. These networks collectively sample just a few square kilometres in the Amazon and Africa; they have no data from the large and diverse tropical forests of southeast Asia.

Beyond determining the size and location of the forest sink, scientists are trying to assess whether it is changing. In March, the RAINFOR team analysed more than 850,000 measurements of approximately 189,000 individual trees and found that the large Amazon forest carbon sink seems to be shrinking<sup>5</sup>. Carbon uptake in their plots during the past decade was one-third smaller than during the 1990s.

**“I THINK OF IT LIKE THESE PERFORMANCE ENHANCERS THAT SOME STELLAR ATHLETES USE”.**

The researchers suspect multiple factors might be at play. Major droughts that hit the Amazon in 2005 and 2010 could have slowed tree growth during this period. Meanwhile, rising temperatures and CO<sub>2</sub> levels may be accelerating the life cycles of trees: if so, trees are now dying earlier than expected, says Roel Brien, an ecologist at the University of Leeds, UK, and lead author of the study.

Some other researchers are not convinced by the evidence. Helene Muller-Landau, an ecologist at the Smithsonian’s Tropical Forest Research Institute on Barro Colorado Island in Panama, thinks that the RAINFOR group is finding an apparent decline now largely because it overestimated the Amazonian carbon sink during the 1990s. The group’s plots, she says, sample too small an area — just three square kilometres out of the vast two-million-square-kilometre Amazon — to support its broad claims. “If you actually look at the area covered, it’s just so pitifully small,” she says.

There can also be bias in how researchers have typically chosen plots and measured biomass, Muller-Landau says. Tropical forests can be hot, humid, buggy, dangerous and in some cases nearly impossible to reach. So rather than sample randomly, scientists often choose study sites based on ease of access. And biomass estimates vary depending on the choice of species-specific equations used to convert circumference and height measurements; for many tropical trees, reliable equations are still being worked out.

Although no one doubts that forests are taking up some of the CO<sub>2</sub> emitted by human activity, scientists are still unsure which forests are sequestering the most carbon, and how much is stored in long-lasting wood versus in roots and soil.

## HELP FROM ABOVE

Researchers will only ever be able to measure a tiny fraction of the world’s trees by wrapping tapes around them one at a time, so they are taking to the skies to get a broader perspective. Some planes and satellites are outfitted with laser-based lidar systems that measure the height of the tree tops. Scientists can then estimate an area’s biomass by using the forest’s average canopy height and tree type.

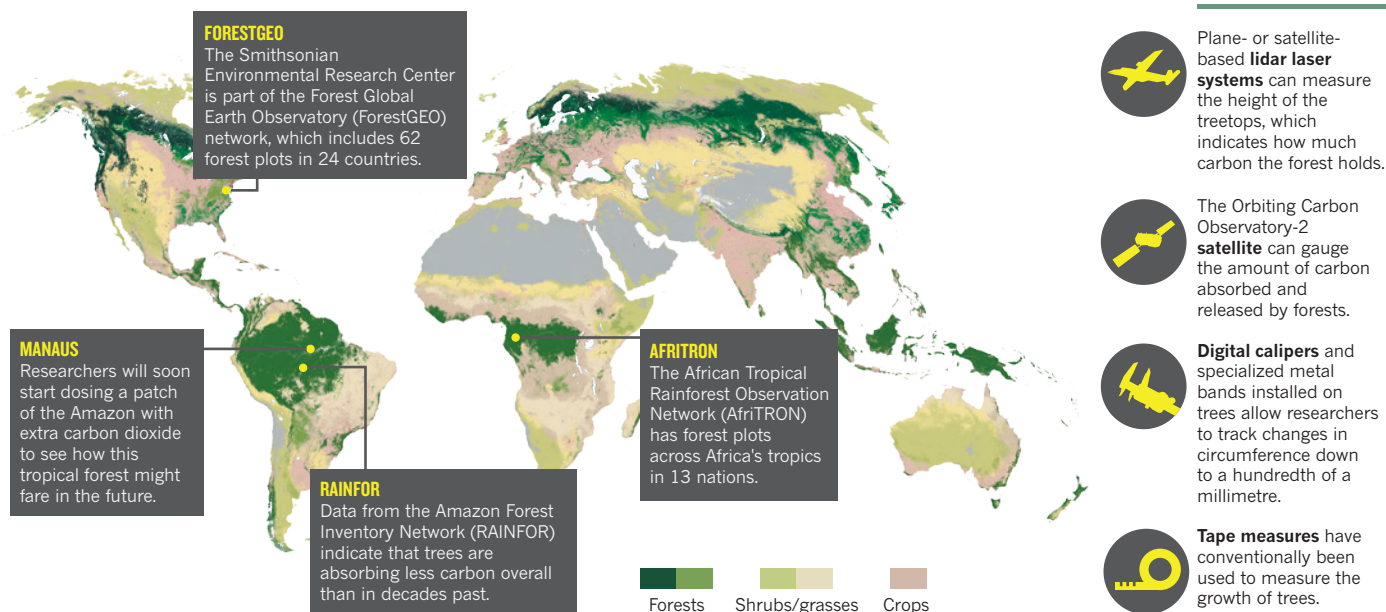
Plane-mounted lidar can collect data for 35,000 hectares in one hour, says Gregory Asner, an ecologist with the Carnegie Institution for Science in Stanford, California. The uncertainties in his lidar-based forest biomass estimates are now down to around 10%, comparable to those from ground-based studies, he says, although others say the uncertainties in both types of estimates are larger.

For a truly global view, scientists agree that nothing beats a satellite. Current Earth-observing satellites lack the resolution of plane- or ground-based measurements but can fill in areas where data are scarce or non-existent. NASA’s Orbiting Carbon Observatory-2 (OCO-2),



# TREE TALES

Networks of research sites around the globe indicate that forests absorb and store about one-quarter of the roughly 10 billion tonnes of carbon emitted by burning fossil fuels each year. But the size of that carbon sink may be shrinking.



launched in July 2014, will soon provide fresh data to help locate the missing sink. The satellite uses spectrometers to measure concentrations of CO<sub>2</sub> to within a few parts per million, allowing scientists to pinpoint the locations where carbon is being emitted and sequestered (see “Tree tales”). A separate instrument can determine how much photosynthesis is occurring at a specific location. Although OCO-2 does not measure tree biomass directly, it will provide enough data for scientists to determine how much carbon is entering and leaving different ecosystems. NASA expects to release preliminary results from the satellite by the end of the year, but it will be at least several years before the data can address whether forest sinks are changing. And even then, the OCO-2 measurements won’t answer whether carbon is going into trees, soil or somewhere else, so ground-based observations will still be needed, says David Crisp, chief scientist for OCO-2.

## TOMORROW'S TREES

Other scientists seeking to predict the carbon sink's future are turning the clocks forward — with experiments that expose today's forests to future conditions. One strategy involves piping CO<sub>2</sub> into a forest to raise concentrations from the present 400 parts per million to roughly 550 parts per million — a level expected before this century's end.

In experiments in the United States and Europe, trees dosed with extra CO<sub>2</sub> grew faster, just as expected. But the effects often did not last. One explanation is that enhanced trees may quickly use up other vital nutrients, such as nitrogen, says ecologist Richard Norby at Oak Ridge National Laboratory in Tennessee, who led one of the experiments.

Researchers from the United States, the United Kingdom and Brazil are now building a CO<sub>2</sub>-enrichment experiment near Manaus, Brazil (see *Nature* **496**, 405–406; 2013), which they hope to start next year. That experiment will provide valuable information about trees in the tropics, but it will not represent the future of all forests in that region, says Simon Lewis, an ecologist at the University of Leeds and University College London who is involved in the RAINFOR and AfriTRON networks. The region around Manaus has poorer soils than other parts of the Amazon so trees grow more slowly, says Lewis, and “it will take longer for the impacts to be seen”.

In the meantime, researchers are trying other methods to peer into the future. Some 20 teams have built Earth-system models that seek to simulate the climate and vegetation on the planet, including how carbon

moves between the oceans, atmosphere and continents. These models currently represent forests in a simplified manner, and they disagree about the future. Some predict that forests will continue to soak up massive amounts of carbon in coming decades, whereas others suggest that forests could become stressed by droughts and high temperatures and die back, releasing carbon into the atmosphere.

The emerging insights about forests — from individual tree measurements to satellite data to computer simulations — will all play a part in how countries decide to manage their resources. And that has implications for global climate negotiations because some carbon-reduction schemes rely on rewarding nations for keeping carbon locked up in forests. For that to work, researchers will need to find reliable ways to track the changing amounts of forest carbon. The current level of uncertainty in forest biomass estimates “does not exactly provide a lot of confidence”, Muller-Landau says. “Having something verifiable would have to be fairly key” for carbon accounting, she adds.

To that end, scientists such as Parker are developing more precise ways to monitor trees growing in their experimental plots. On a cloudy spring day at the Smithsonian's Chesapeake site, Parker directs volunteers to install spring-tensioned steel bands called dendrometers in a 130-year-old stand. As the tree trunks expand over time, they will widen gaps in the bands, which can be measured using digital calipers. The technique can track changes down to a hundredth of a millimetre — thinner than a human hair — giving researchers an unprecedented ability to study growth patterns. The method can even detect how trees swell and contract over a few hours as they absorb or lose water.

By the end of the day, Parker's team has finished attaching several more dendrometers. More than a thousand trees at the Smithsonian centre now sport the metal rings, and their number is increasing around the world. Parker puts his equipment in a truck and drives off towards home. But he and his crew will be back soon to check how their trees are responding as Earth's climate — and its forests — enter uncharted territory. ■

**Gabriel Popkin** is a freelance writer in Mount Rainier, Maryland.

1. Odum, E. P. *Science* **164**, 262–270 (1969).
2. McMahon, S. M., Parker, G. G. & Miller, D. R. *Proc. Natl Acad. Sci. USA* **107**, 3611–3615 (2010).
3. Keenan, T. F. *et al. Nature* **499**, 324–327 (2013).
4. Pan, Y. *et al. Science* **333**, 988–993 (2011).
5. Brienen, R. J. W. *et al. Nature* **519**, 344–348 (2015).



# THE ROBOT'S DILEMMA

Working out how to build ethical robots is one of the thorniest challenges in artificial intelligence.

BY BOER DENG



In his 1942 short story ‘Runaround’, science-fiction writer Isaac Asimov introduced the Three Laws of Robotics — engineering safeguards and built-in ethical principles that he would go on to use in dozens of stories and novels. They were: 1) A robot may not injure a human being or, through inaction, allow a human being to come to harm; 2) A robot must obey the orders given it by human beings, except where such orders would conflict with the First Law; and 3) A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.

Fittingly, ‘Runaround’ is set in 2015. Real-life roboticists are citing Asimov’s laws a lot these days: their creations are becoming autonomous enough to need that kind of guidance. In May, a panel talk on driverless cars at the Brookings Institution, a think tank in Washington DC, turned into a discussion about how autonomous vehicles would behave in a crisis. What if a vehicle’s efforts to save its own passengers by, say, slamming on the brakes risked a pile-up with the vehicles behind it? Or what if an autonomous car swerved to avoid a child, but risked hitting someone else nearby?

“We see more and more autonomous or automated systems in our daily life,” said panel participant Karl-Josef Kuhn, an engineer with Siemens in Munich, Germany. But, he asked, how can researchers equip a robot to react when it is “making the decision between two bad choices”?

The pace of development is such that these difficulties will soon affect health-care robots, military drones and other autonomous devices capable of making decisions that could help or harm humans. Researchers are increasingly convinced that society’s acceptance of such machines will depend on whether they can be programmed to act in ways that maximize safety, fit in with social norms and encourage trust. “We need some serious progress to figure out what’s relevant for artificial intelligence to reason successfully in ethical situations,” says Marcello Guarini, a philosopher at the University of Windsor in Canada.

Several projects are tackling this challenge, including initiatives funded by the US Office of Naval Research and the UK government’s engineering-funding council. They must address tough scientific questions, such as what kind of intelligence, and how much, is needed for ethical decision-making, and how that can be translated into instructions for a machine. Computer scientists, roboticists, ethicists and philosophers are all pitching in.

“If you had asked me five years ago whether we could make ethical robots, I would have said no,” says Alan Winfield, a roboticist at the Bristol Robotics Laboratory, UK. “Now I don’t think it’s such a crazy idea.”

## LEARNING MACHINES

In one frequently cited experiment, a commercial toy robot called Nao was programmed to remind people to take medicine.

“On the face of it, this sounds simple,” says Susan Leigh Anderson, a philosopher at the University of Connecticut in Stamford who did the work with her husband, computer scientist Michael Anderson of the University of Hartford in Connecticut. “But even in this kind of limited task, there are nontrivial ethics questions involved.” For example, how should Nao proceed if a patient refuses her medication? Allowing her to skip a dose could cause harm. But insisting that she take it would impinge on her autonomy.

To teach Nao to navigate such quandaries, the Andersons gave it examples of cases in which bioethicists had resolved conflicts involving autonomy, harm and benefit to a patient. Learning algorithms then sorted through the cases until they found patterns that could guide the robot in new situations<sup>1</sup>.

With this kind of ‘machine learning’, a robot can extract useful knowledge even from ambiguous inputs (see [go.nature.com/2r7nav](http://go.nature.com/2r7nav)). The approach would, in theory, help the robot to get better at ethical

decision-making as it encounters more situations. But many fear that the advantages come at a price. The principles that emerge are not written into the computer code, so “you have no way of knowing why a program could come up with a particular rule telling it something is ethically ‘correct’ or not”, says Jerry Kaplan, who teaches artificial intelligence and ethics at Stanford University in California.

Getting around this problem calls for a different tactic, many engineers say; most are attempting it by creating programs with explicitly formulated rules, rather than asking a robot to derive its own. Last year, Winfield published the results<sup>2</sup> of an experiment that asked: what is the simplest set of rules that would allow a machine to rescue someone in danger of falling into a hole? Most obviously, Winfield realized, the robot needed the ability to sense its surroundings — to recognize the position of the hole and the person, as well as its own position relative to both. But the robot also needed rules allowing it to anticipate the possible effects of its own actions.

## “We need some serious progress to figure out what’s relevant for artificial intelligence to reason successfully in ethical situations.”

Winfield’s experiment used hockey-puck-sized robots moving on a surface. He designated some of them ‘H-robots’ to represent humans, and one — representing the ethical machine — the ‘A-robot’, named after Asimov. Winfield programmed the A-robot with a rule analogous to Asimov’s first law: if it perceived an H-robot in danger of falling into a hole, it must move into the H-robot’s path to save it.

Winfield put the robots through dozens of test runs, and found that the A-robot saved its charge each time. But then, to see what the allow-no-harm rule could accomplish in the face of a moral dilemma, he presented the A-robot with two H-robots wandering into danger simultaneously. Now how would it behave?

The results suggested that even a minimally ethical robot could be useful, says Winfield: the A-robot frequently managed to save one ‘human’, usually by moving first to the one that was slightly closer to it. Sometimes, by moving fast, it even managed to save both. But the experiment also showed the limits of minimalism. In almost half of the trials, the A-robot went into a helpless dither and let both ‘humans’ perish. To fix that would require extra rules about how to make such choices. If one H-robot were an adult and another were a child, for example, which should the A-robot save first? On matters of judgement like these, not even humans always agree. And often, as Kaplan points out, “we don’t know how to codify what the explicit rules should be, and they are necessarily incomplete”.

Advocates argue that the rule-based approach has one major virtue: it is always clear why the machine makes the choice that it does, because its designers set the rules. That is a crucial concern for the US military, for which autonomous systems are a key strategic goal. Whether machines assist soldiers or carry out potentially lethal missions, “the last thing you want is to send an autonomous robot on a military mission and have it work out what ethical rules it should follow in the middle of things”, says Ronald Arkin, who works on robot ethics software at Georgia Institute of Technology in Atlanta. If a robot had the choice of saving a soldier or going after an enemy combatant, it would be important to know in advance what it would do.

With support from the US defence department, Arkin is designing a program to ensure that a military robot would operate according to international laws of engagement. A set of algorithms called an ethical governor computes whether an action such as shooting a missile

**The fully programmable Nao robot has been used to experiment with machine ethics.**

**➔ NATURE.COM**  
For a podcast on robot ethics, see: [go.nature.com/wvkakj](http://go.nature.com/wvkakj)



'Robear' is designed to help to care for ill or elderly people.

is permissible, and allows it to proceed only if the answer is 'yes'.

In a virtual test of the ethical governor, a simulation of an unmanned autonomous vehicle was given a mission to strike enemy targets — but was not allowed to do so if there were buildings with civilians nearby. Given scenarios that varied the location of the vehicle relative to an attack zone and civilian complexes such as hospitals and residential buildings, the algorithms decided when it would be permissible for the autonomous vehicle to accomplish its mission<sup>3</sup>.

## “Logic is how we reason and come up with our ethical choices.”

Autonomous, militarized robots strike many people as dangerous — and there have been innumerable debates about whether they should be allowed. But Arkin argues that such machines could be better than human soldiers in some situations, if they are programmed never to break rules of combat that humans might flout.

Computer scientists working on rigorously programmed machine ethics today favour code that uses logical statements, such as 'If a statement is true, move forward; if it is false, do not move.' Logic is the ideal choice for encoding machine ethics, argues Luís Moniz Pereira, a computer scientist at the Nova Laboratory for Computer Science and Informatics in Lisbon. “Logic is how we reason and come up with our ethical choices,” he says.

Crafting instructions capable of the logical steps that go into making ethical decisions is a challenge. For example, Pereira notes, the logical languages used by computer programs have trouble coming to conclusions about hypothetical scenarios, but such counterfactuals are crucial in resolving certain ethical dilemmas.

One of these is illustrated by the trolley problem, in which you imagine a runaway railway trolley is about to kill five innocent people who are on the tracks. You can save them only if you pull a lever

that diverts the train onto another track, where it will hit and kill an innocent bystander. What do you do? In another set-up, the only way to stop the trolley is to push the bystander onto the tracks.

People often answer that it is all right to stop the trolley by hitting the lever, but viscerally reject the idea of pushing the bystander. The basic intuition, known to philosophers as the doctrine of double effect, is that deliberately inflicting harm is wrong, even if it leads to good. However, inflicting harm might be acceptable if it is not deliberate, but simply a consequence of doing good — as when the bystander simply happens to be on the tracks.

This is a very difficult line of analysis for a decision-making program. To begin with, the program must be able to see two different futures: one in which a trolley kills five people, and another in which it hits one. The program must then ask whether the action required to save the five is impermissible because it causes harm, or permissible because the harm is only a side effect of causing good.

To find out, the program must be able to tell what would happen if it chose not to push the bystander or pull the lever — to account for counterfactuals. “It would be as if a program was constantly debugging itself,” says Pereira — “finding where in a line of code something could be changed, and predicting what the outcome of the change would be.” Pereira and Ari Saptawijaya, a computer scientist at the University of Indonesia in Depok, have written a logic program<sup>4</sup> that can successfully make a decision based on the doctrine of double effect, as well as the more sophisticated doctrine of triple effect, which takes into account whether the harm caused is the intended result of the action, or simply necessary to it.

### HUMANS, MORALS, MACHINES

How ethical robots are built could have major consequences for the future of robotics, researchers say. Michael Fisher, a computer scientist at the University of Liverpool, UK, thinks that rule-bound systems could be reassuring to the public. “People are going to be scared of robots if they're not sure what it's doing,” he says. “But if we can analyse and prove the reasons for their actions, we are more likely to surmount that trust issue.” He is working with Winfield and others on a government-funded project to verify that the outcomes of ethical machine programs are always knowable.

By contrast, the machine-learning approach promises robots that can learn from experience, which could ultimately make them more flexible and useful than their more rigidly programmed counterparts. Many roboticists say that the best way forward will be a combination of approaches. “It's a bit like psychotherapy,” says Pereira. “You probably don't just use one theory.” The challenge — still unresolved — is to combine the approaches in a workable way.

These issues may very soon come up in the fast-moving field of autonomous transport. Already, Google's driverless cars are zipping across parts of California (see *Nature* **518**, 20–23; 2015). In May, autonomous trucks from German car-maker Daimler began driving themselves across the Nevada desert. Engineers are thinking hard about how to program cars to both obey rules and adapt to situations on the road. “Up until now we've been trying to do things with robots that humans are bad at,” such as maintaining attention on long drives or being quick on the brakes when the unexpected occurs, says Bernhard Weidemann, a spokesperson for Daimler in Stuttgart. “Going forward, we will have to try to program things that come more naturally to humans, but not to machines.” ■

**Boer Deng** is a news intern for *Nature* in Washington DC.

1. Anderson, M. & Anderson, S. L. *AI Magazine* **28**, 15–26 (2007).
2. Winfield, A. F. T., Blum, C. & Liu, W. in *Advances in Autonomous Robotics Systems* 85–96 (Springer, 2014).
3. Arkin, R. C., Ulam, P. & Duncan, B. *An Ethical Governor for Constraining Lethal Action in an Autonomous System* Technical Report GIT-GVU-09-02 (2009).
4. Pereira, L. M. & Saptawijaya, A. in *Logic, Argumentation and Reasoning* (eds Urbaniak, R. & Payette, G.) (Springer, in the press); available at <http://go.nature.com/3xlske>



# COMMENT

**POLICY** There must be more to the future of transport than the driverless car **p.29**

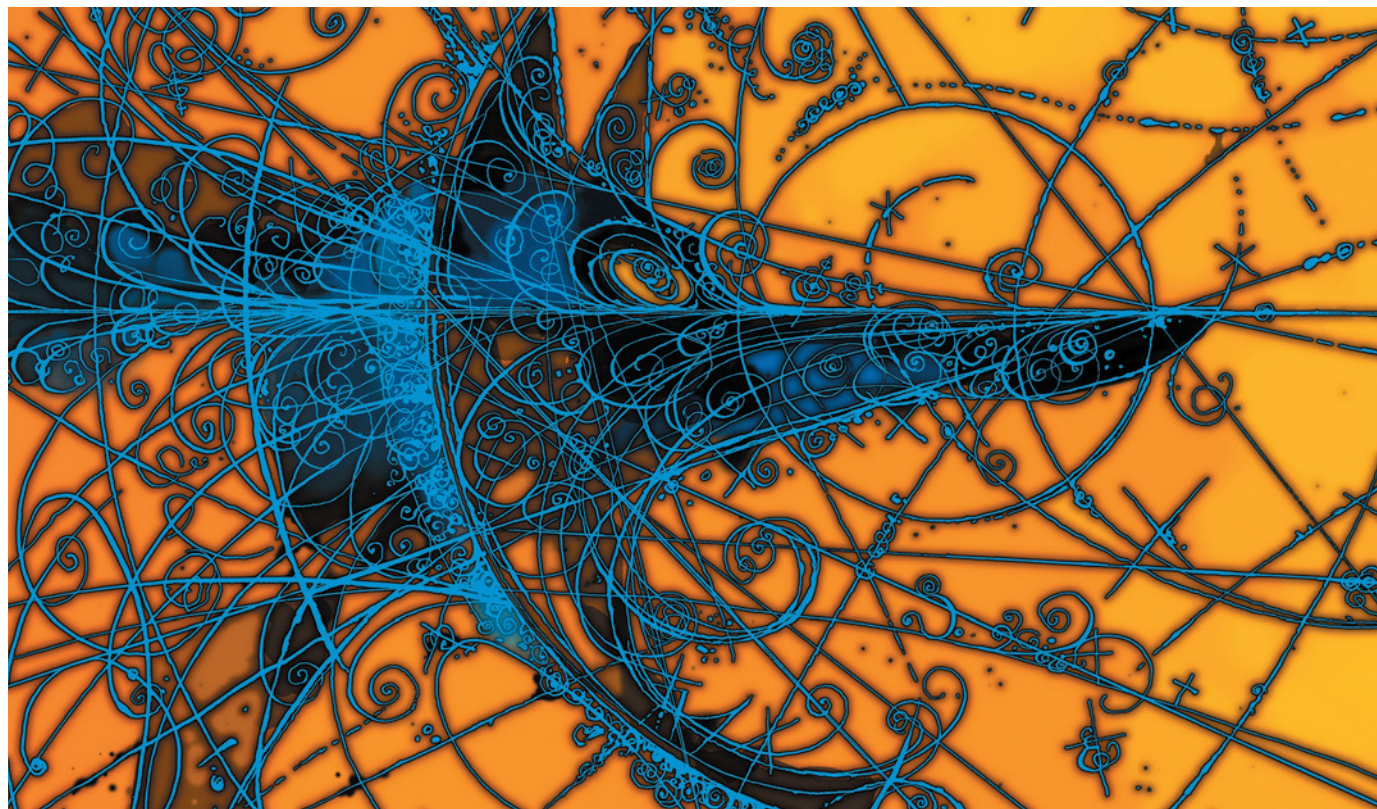
**ENERGY** Clarity needed about power output of renewables per unit area **p.32**

**PALAEONTOLOGY** Who is hominin discovery centre for? **p.33**



**IMMUNOLOGY** A humanized history of monoclonal antibodies **p.34**

ILLUSTRATION BY DAVID PARKINS



## Don't cry wolf

Tighten the requirements for declaring physics breakthroughs, says **Jan Conrad**.

The past few years have seen a slew of announcements of major discoveries in particle astrophysics and cosmology. The list includes faster-than-light neutrinos; dark-matter particles producing  $\gamma$ -rays; X-rays scattering off nuclei underground; and even evidence in the cosmic microwave background for gravitational waves caused by the rapid inflation of the early Universe. Most of these turned out to be false alarms; and in my view, that is the probable fate of the rest.

There are consequences to broadcasting seemingly extraordinary results to peers and the public before they are reviewed, or despite knowing that better data are just around the corner. Colleagues who once got excited now shake their heads and joke about 'yet another dark-matter candidate'. The field

has cried wolf too many times and lost credibility. One colleague told me that granting panels are becoming wary of funding astrophysical searches for dark-matter particles.

I also worry that false discoveries are undermining public trust in science. As cosmic phenomena come and go — not to mention endless speculation about hypothetical concepts such as parallel and holographic universes — why should anyone believe that any scientific result will hold?<sup>1</sup>

Several trends have brought us to this state of affairs. Intense competition, increased use of public data sets and online publishing of draft papers without proper refereeing have eroded traditional standards for making extraordinary claims.

Particle physics and astrophysics pioneered

the open release of data and publications more than two decades ago; other disciplines are following their lead. The scientific community must now address the habits that have crept in to ensure that enticing reports of false discoveries do not overwhelm more sober accounts of genuine scientific breakthroughs.

### SHIFTING PRACTICES

Three changes in the ways that scientific studies are done and reported are fuelling this rash of false discoveries.

First, statistical standards have fallen. Extraordinary claims demand extraordinary proof. In particle physics, the usual threshold is '5 sigma': a signal 5 times stronger than the average noise level (sigma), which translates to a roughly 1-in-3.5-million probability ▶

► that the results were due to chance. But 5-sigma claims are becoming rare as scientists rush to assert priority with exciting but tentative results. The July 2012 official announcement of the discovery of the Higgs boson with the Large Hadron Collider at CERN, Europe's particle-physics lab near Geneva, Switzerland, was preceded by press releases of weak but suggestive indications even though there was no competition.

That scientists change the wording in their papers from 'discovery' to 'evidence' or 'indication' has little influence on how the results are used. Take the latest dark-matter discovery claim. On 8 March, astronomers posted a preprint of a paper in the arXiv repository, and their university issued a press release reporting what the authors called a "tantalizing" sign of  $\gamma$ -rays coming from a recently found dwarf-galaxy companion to the Milky Way that is allegedly rich in dark matter<sup>2</sup>. The  $\gamma$ -ray signal, found in images from the Fermi  $\gamma$ -ray satellite's Large Area Telescope (LAT), seemed to be consistent with high-energy radiation produced when particles of dark matter annihilate. But the photon excess of only 3–4 times the noise level was inconclusive, as the authors acknowledged.

Another paper posted on arXiv the same day disfavoured the discovery. A more comprehensive re-analysis of the same data by the Fermi-LAT instrument team<sup>3</sup> — using updated software 30–40% more sensitive — recorded no signal beyond noise. The authors of the first paper acknowledged that the software upgrade was imminent and would confirm or refute their claim, but did not wait for it.

Detecting a noise fluctuation is nothing new, but the possibility that the 'detection' might have been dark matter meant that it was widely reported in the media. Even balanced reporting raises the issue in the public's mind; the account in *The New York Times*<sup>4</sup> mentioned the non-detection, but the hint of excitement drove the story.

Second, the greater use of public data sets increases the risk that some researchers will make spurious detections near the edge of an instrument's sensitivity. More brains may be picked to mine the data. But analysis is difficult without inside information from those who built and calibrated the instrument.

That was the case with the Fermi-LAT dark-matter detection. The released Fermi-LAT data — public since 2009 — are the product of complicated algorithms and calibrations that turn the electronic signals of detectors into quantities that any physicist can in principle analyse. The instrument builders, however, have the know-how to push the noise limits down.

The risk that someone will misuse data also grows when more people have access to them. Even the largest collaborations cannot police discoveries made by outsiders using

their data. Even if they internally re-run the analysis, the damage is done once an erroneous result has been made public.

Third, many more papers are now released on preprint servers such as arXiv (which had about 100,000 submissions in 2014), and press releases are sent out before peer review. Competition for positions, funding, career metrics such as the  $h$ -index and prizes drives

**"Metrics need to be devised that distinguish citations of discredited claims."**

the rush to publish prematurely and publicize results.

Incorrect papers posted on arXiv do more than add to the noise of irrelevant results. Funding decisions are skewed; theorists waste a lot of time trying to devise explanations; and the public is misled through news reports.

A striking example of a premature claim released online before peer review was the report last year of evidence for gravitational waves and cosmic inflation — the Universe's rapid expansion in the instant after the Big Bang — by the BICEP2 microwave telescope at the South Pole. The detection of a swirled polarization pattern, known as a B-mode, in the cosmic microwave background (radiation left over from the Big Bang) was not in doubt — it had a 7-sigma signal<sup>5</sup>. But its supposed cosmic origin turned out to be false. It was shown six months later — with data from the European Space Agency's Planck satellite — to be warm dust in the Milky Way<sup>6</sup>. Again, I believe that the authors of the original paper must have known of the impending Planck data but chose to blow their trumpet ahead of confirmation. Eventually, the BICEP2 and Planck collaborations worked together to arrive at a solid result<sup>7</sup>, an approach that should have been considered from the beginning.

#### QUALITY CONTROL

To avoid further weakening of scientific standards and reputations, researchers need to stick to scientific best practice.

A first step would be for physicists to make sure that they apply the 5-sigma rule (or an equivalent) for firm discoveries. Online posting should not be elided with publication. It is premature to announce an important finding to the public at the same time as it is announced to scientific peers. Critical examination by peers is necessary — not least to avoid personal biases.

As long as online posting is confused with the release of deeply scrutinized results, quality assurance of preprints posted online should become stricter. An 'endorsement system', whereby users must be endorsed by other users before posting a paper, has been developed by arXiv to ensure that non-scientific pieces are not hosted there. More is needed for extraordinary claims. Named

reviewers for major discoveries would reassure the readers and authors, as well as crediting the reviewers. Journals should discourage the referencing of arXiv papers.

Instrument builders and specialists who collected the original data should review major claims that are based on public data sets, either as referees, advisers or collaborators. Other teams with ancillary data that could refute or prove a claim should be involved in checking major results before release. This will require voluntary good conduct by competitors, which again could be encouraged by naming reviewers on breakthrough papers.

A system needs to be established to reward best practice. Collaborations should establish a way to ensure that a data team working with an individual scientist will not competitively sink the scientist's publication nor diminish their visibility. Internal review should precede announcements of major results at conferences. Policies should be devised for author lists to give proper credit.

Journals and arXiv should find a strategy for allocating credit to the lead scientists in such collaborations. The BICEP2 team, for example, did work with the Planck collaboration later; had they been able to mark their priority better they might have delayed a press conference.

Not surprisingly, the original BICEP2 paper has ten times more citations than the final word; many incorrect papers are more highly cited than counter cases. Academic metrics need to be devised that distinguish citations of discredited claims so that it is not more advantageous to state and retract a result than to make a solid discovery.

Physicists' associations (such as the American Physical Society or the International Union of Pure and Applied Physics) should lead a movement akin to the biology community's reproducibility initiatives. Scientists, publishers and representatives of funding agencies must convene to discuss improvements to norms such as peer review, metrics, use of databases, quality assurance and codes of conduct. ■

**Jan Conrad** is professor of astroparticle physics at the Oskar Klein Centre for Cosmoparticle Physics, Stockholm University, Stockholm, Sweden.  
e-mail: conrad@fysik.su.se

1. Ellis, G. & Silk, J. *Nature* **516**, 321–323 (2014).
2. Gerlinger-Sameth, A. et al. Preprint at <http://arxiv.org/abs/1503.02320> (2015).
3. Fermi LAT Collaboration & DES Collaboration. Preprint at <http://arxiv.org/abs/1503.02632> (2015).
4. Overbye, D. *The New York Times* 'Gamma Rays May Be Clue on Dark Matter' (11 March 2015).
5. Ade, P. A. R. et al. *Phys. Rev. Lett.* **112**, 241101 (2014).
6. Planck Collaboration. *Astron. Astrophys.* **576**, A107 (2015).
7. Ade, P. A. R. et al. *Phys. Rev. Lett.* **114**, 101301 (2015).





Cities such as Lahore in Pakistan can have traffic jams that last for hours.

# Six research routes to steer transport policy

Strategies must better balance the costs and benefits of travel and be realistic about the promises of new technologies, say **Eric Bruun** and **Moshe Givoni**.

Society is not balancing the benefits and costs of travel. Although technology enables us to get around faster than ever, many cities are gridlocked — São Paulo in Brazil frequently experiences eight-hour traffic jams. More than 90% of the 1.2 million traffic deaths each year worldwide occur in developing countries and half involve pedestrians, cyclists and motorcyclists. Premature death from vehicle-related fine-particle air pollution worldwide is predicted to rise by 50% by 2030. In rich countries, sedentary lifestyles and obesity are in large part the result of our love affair with the car. In poor countries, people may spend two hours walking to work to avoid a modest bus fare.

Transport research is central to twenty-first-century global challenges that include energy provision, climate change and health. Yet the field is stuck. The language is changing — ‘transport’ has become ‘mobility’ — and sustainability is more often mentioned

in research papers and policy documents. But most planners are still hopelessly trying to fight congestion, and most researchers and policy-makers put too much faith in technological solutions.

Reframing mobility research to answer the following six questions will inform better transport policies.

## SIX QUESTIONS

**What are the long-term impacts of new technologies?** Although the excitement associated with a new product, service or tool is often justified, the negative, unintended impacts must be anticipated.

Take the driverless car. Depending on whom one asks, such cars will be in wide use in some countries by 2025 or 2050. They are framed as a technology that offers cheap mobility while saving time and energy<sup>1</sup>. But it was exactly this thinking that brought us the ‘with-driver’ private car and its

unsustainable consequences.

The driverless car promises to be even more successful. Getting people out of their driverless cars will be even harder.

On average, people around the world spend an hour a day travelling, a pattern that has held for centuries and across cultures. When we are able to eat, sleep and work in our driverless cars, this time will become longer, creating a burst of urban sprawl with its associated increases in energy consumption and adverse impacts on the land.

The stakes are too high to believe the promises of new mobility technologies without extensive research that goes beyond the technical, regulatory and commercial. Researchers and policy-makers need to treat any significant technological change as a ‘socio-technical’ change that alters daily practices and functioning. Protection of personal privacy will be a particular challenge. On the basis of impact assessments,

governments may need to discourage certain new technologies or encourage their utilization in a particular way. For example, driverless vehicles hold great potential for public transportation.

**How should the impacts of transport systems be evaluated?** Economic cost-benefit analysis is an increasingly controversial method for assessing investment and policy decisions. Transport affects so many aspects of life, particularly in urban areas, over such a long time that a monetary focus alone cannot do the issue justice. Analysis of multiple criteria offers some improvement on cost-benefit, but it is unrealistic to expect to capture all impacts in one score.

Who benefits and who doesn't needs to be accounted for. For example, building motorways through US cities in the 1960s divided low-income black and minority ethnic communities while enabling 'white flight' to the suburbs. Moreover, under the current accounting system, future generations lose out as the discounting of costs and benefits in the future encourages the consumption of non-renewable resources now.

The value placed on travel time needs to be reconsidered. It can be a waste and viewed as a monetary cost, but with wireless technology or a good book it can also be productive and fun — even more so when the driverless car arrives.

Travel behaviour models — used to project future demand — are crucial for any evaluation. Increasingly sophisticated models that are largely based on random utility theory from mathematical psychology have been developed over the past 30 years to better capture reality. But they are so complex and expensive that most cities cannot afford them or collect the data required. Results that can be comprehended only by the modellers are not transparent enough to support democratic decision-making.

Researchers must come up with new

evaluation methods that are robust and scientifically defensible. The outputs must be comprehensible to elected officials and to the public. Such methods must include both quantitative and qualitative benefits and costs, and capture a much larger array of them. For example, researchers might consider whether the public perceives that the comfort and beauty of their city will improve or deteriorate after a major investment.

A good example is the Øresund Eco-Mobility project, a joint effort between Swedish and Danish city and regional governments and universities to rethink transport in cities around the Øresund strait, including Copenhagen and Malmö. It is combining cost-benefit analysis with risk assessment and qualitative impacts of interest to the population, determined at a 'decision conference' at which planners and public representatives can see and hear differences of opinion.

***"The majority of research money for transport currently goes to technological development with commercial potential."***

**How does the structure of cities affect sustainability, living standards and functioning costs?** Studies of urban density, energy consumption and travel statistics show that altering the form of cities to reduce greenhouse-gas emissions is as effective as improving technologies and substituting fuels, yet it receives much less attention<sup>2</sup>. Energy consumption per capita by private vehicles declines with higher urban density, for instance (see 'Transport trends')<sup>3</sup>.

Total expenditures (public and private) on passenger transport decrease as urban density increases. Yet zoning and infrastructure investment decisions are not based on broader scientific analyses of the impacts. Understanding the drivers of sprawl is of vital importance for fast-growing cities such as Mexico City, Delhi or Lahore in Pakistan that are swallowing up adjacent farmland and wetlands.

Researchers need to consider more fully urban transport alongside other features of the built environment. We know, for example, that taller buildings and smaller areas used per person for a given type of activity (residential, commercial or recreational) tend to be associated with compact and energy-efficient cities, such as Hong Kong. Yet planners lack accurate models for land use development that consider many design variables.

Scientists and planners urgently need to understand the significance of changes in urban development plans on the construction and operating costs of all aspects of the built environment, on total urban energy consumption, on living standards and on space consumption. To accomplish this,

universities and urban governments need to break down traditional borders between disciplines and professional responsibilities.

**How can mobility beyond cities be improved?** Transport links beyond cities are important for regional development. The exodus of people from the countryside to cities in search of employment is intensifying urbanization worldwide and increasing pressure on scarce resources. Urbanization scholars highlight the rapid growth of megacities, but smaller cities and villages must also be considered.

Researchers need to model how better connections — local, regional and between a city and its hinterlands — might improve the prospects of smaller towns and rural areas. Studies need to be more inclusive, politically neutral and regionally equitable. Transport investments often favour large cities and their links with expensive, fast transport options. The opportunity costs — what a government could have done with the money — are not considered sufficiently.

For example, the UK government plans to spend more than £40 billion (US\$63 billion) on High Speed 2, a 400-kilometre-per-hour rail link between London and Birmingham and (later) Manchester. In our view, the money would be better spent on improving the country's entire public-transport network, which is poor by European standards, and on local and regional transport, which is currently dominated by private-car use<sup>4</sup>. Joining up many small cities could benefit the national economy and society: for every 12 jobs created in cities in the south of England between 2004 and 2013, only one was created elsewhere in Britain. Many other countries and regions face similar decisions.

**How could transport be improved in developing countries?** Researchers need to assess and suggest ways to establish rapid, cheap and effective transport systems in poor nations. Elaborate physical redesigns of infrastructure, similar to those made by high-income countries, take too much time and money to implement. Instead, developing countries should learn from developed countries' planning mistakes. They could also 'leapfrog' to the latest technologies.

For example, in Nairobi in 2013, student bus passengers were issued with smartphones that allowed city planners and researchers to track their routes, count riders and identify areas of congestion (see [go.nature.com/ihhy6t](http://go.nature.com/ihhy6t)). An app that contacts a traffic signal to let a bus through an intersection can aid mobility flow. Private-sector investment in such systems is low because of the lack of commercial prospects for what seems like simple technology. Public-sector funding of such applied research and collaborations between universities from high- and



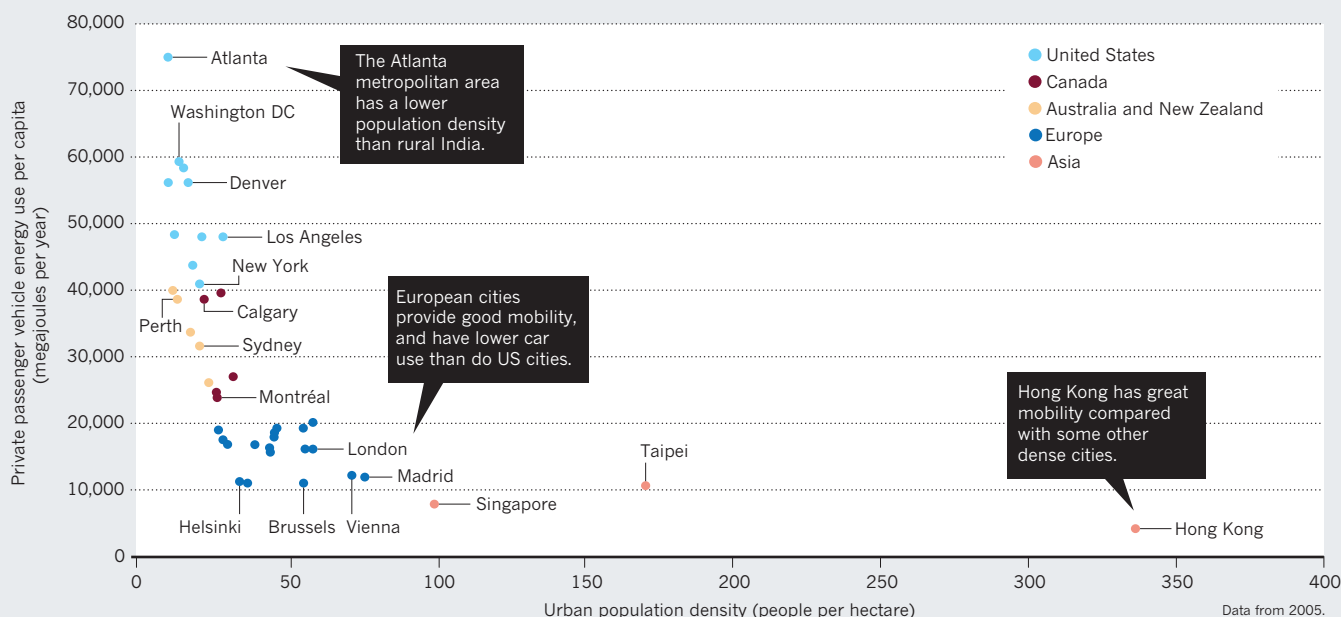
New York City is introducing more bike lanes.

TOMAS ABAD/ALAMY



## TRANSPORT TRENDS

Sprawling metropolises such as Atlanta, Georgia, have higher energy consumption per capita because their residents are highly dependent on cars for long journeys, whereas cities such as London and Hong Kong have higher building densities and more public transport.



low-income countries should be encouraged.

But technology can get in the way. For example, farmers in South Asia experience impassable roads during the rainy or monsoon season. Money would be better spent on gravelling or paving farm tracks than on widening major roads linking large cities. Similarly, investing in standard bus services that have been allowed to wither, rather than in more sophisticated high-capacity rail and bus mass-transit systems that mimic those in developed countries, will be more beneficial. It will be cheaper, realized sooner, carries lower risk, and importantly the benefits will be wider spread across the city and its inhabitants.

**What kinds of governance work for the transport system?** Services such as Uber — the taxi smartphone app that connects passengers with drivers in dozens of cities worldwide — and variants of car sharing have caught the transport planning sector by surprise; institutions are not sure whether to support or fight these advances. Like any innovation they are a great opportunity but also carry risks. Even with shared cars, it is physically impossible for large cities to meet everyone's travel needs with what is essentially a variation of single-occupant vehicles. When finding that parking a car is not an issue any more, people may flock again to cars, reversing the accessibility, sustainability and livability trends experienced in many cities such as New York, which is promoting public transport, cycling and walking as modes of transit. So, good public transit will be needed more than ever to compete with the car.

Research is needed to understand the policy implications of rapid changes in transport technologies and systems and how institutions should evolve to accommodate them<sup>5</sup>. Methodologies and tools are needed for devising effective policies, combining them strategically and overcoming implementation barriers such as public and political acceptability.

One methodology that could be expanded is 'policy packaging'<sup>6</sup>, in which a combination of instruments is implemented while steps are taken to minimize unintended effects and increase the chances of an intervention's success. For example, London's congestion-charge scheme — which charges car users a fee to enter the city centre during certain hours — was accompanied by improvements to public transport and heavy discounts for residents within the zone<sup>7</sup>. Such policy packages have been suggested for promoting car-sharing services in European cities (see, for example, [www.spreeproject.com](http://www.spreeproject.com)).

### FRESH THINKING

Governments should support system-level research that is needed by the public sector yet attracts scant funding from the private sector. The majority of research money for transport currently goes to technological development with commercial potential — such as the driverless car — which already receives private funding.

Universities and governments need to realign research incentives to support the interdisciplinary scholarship needed. This includes stable funding and centres that can attract and nurture a variety of talent and

cross-border collaboration, especially where there is lack of commercial potential but great promise to society. If not, researchers will remain in narrow specialities in which funding and publishing are safer.

Our transport systems, as well as our cities, must be planned for people — not for a particular mode of transport or by a handful of companies with vast lobbying power. Delivering low-carbon mobility for all will take fresh thinking<sup>8</sup>. ■

**Eric Bruun** is a visiting professor in the Department of Civil and Environmental Engineering, Aalto University, Espoo, Finland. **Moshe Givoni** is head of the Transport Research Unit, Department of Geography and the Human Environment, Tel Aviv University, Tel Aviv, Israel. e-mails: [eric.bruun@gmx.com](mailto:eric.bruun@gmx.com); [givonim@post.tau.ac.il](mailto:givonim@post.tau.ac.il)

1. Burns, L. D. *Nature* **497**, 181–182 (2013).
2. Bruun, E. C. *Characterizing and Influencing Modal Relationships, Better Public Transit Systems: Analyzing Investments and Performance* 2nd edn Ch. 5 (Routledge, 2014).
3. Newman, P. & Kenworthy, J. *The End of Automobile Dependence: How Cities are Moving Beyond Car-Based Planning* (Island Press, 2015).
4. Givoni, M. & Banister, D. J. *Transp. Geogr.* **22**, 306–307 (2012).
5. Schiller, P. L., Bruun, E. C. & Kenworthy, J. R. *An Introduction to Sustainable Transportation: Policy Planning and Implementation* Ch. 8 (Earthscan/Routledge, 2010).
6. Givoni M., Macmillen, J., Banister D. & Feitelson, E. *Transp. Rev.* **33**, 1–20 (2013).
7. Givoni, M. *Urban Studies* **49**, 1089–1105 (2012).
8. Banister, D., Givoni, M., Macmillen, J. & Schwanen, T. in *Moving Towards Low Carbon Mobility* (eds Givoni, M. & Banister, D.) (Edward-Elgar, 2013).



Wind turbines are low-density sources of power.



**Power Density:**  
A Key to  
Understanding  
Energy Sources  
and Uses  
VACLAV SMIL  
MIT Press: 2015.

TOM PAIVA PHOTOGRAPHY/GETTY

## ENERGY

# Profiles of power

**Arnulf Grubler** examines a study of power output and spatial area — a key concept in discussing renewables.

**V**aclav Smil is a prolific, sometimes controversial, but invariably thought-provoking author. His latest book centres on a simple but important concept. Power density — defined in the book as “energy’s rate of flow [transfer] per unit of surface area” of land or water — matters because these densities differ vastly for different methods of energy generation and use. That difference needs to be reconciled: massive, costly energy infrastructures such as long-range transport and storage are involved.

Megacities, with their concentrations of high-rise buildings, demand enormous quantities of energy in a comparatively small amount of space at any one time — and so have a high density of power demand. But diffuse renewable energy sources such as sunlight or biomass have low energy yields per hectare and intermittent availability, and

so are low-density power supplies. Hence, in any transition towards renewables, cities will require vast renewable-energy hinterlands along with extensive cropland for food.

Smil excels when discussing the historical context of how our economy, cities and industry rose by harnessing high-power-density fossil fuels. He is also good at explaining the fundamentals that underlie the low power densities of renewables, such as the low efficiency by which plants convert the radiant energy of sunlight into chemical energy in the form of biomass (0.5%, Smil shows). Photos illustrate the energy options well.

However, Smil’s technical discussion of options from renewables such as solar and wind to fossil fuels such as coal and gas, with its many site-specific examples, is necessarily repetitive and thus a tedious read. Moreover, he fails to clarify some fundamental concepts

or to differentiate appropriately between different qualitative aspects of power and area.

Most fundamentally, he does not explain the concepts of power and energy per se; he seems to presume that readers are familiar with them. And it is unclear why he focuses on power density rather than energy density, or does

not combine the two (given that energy = power × time). Imagine a stream of water running at 1 litre per second. The total amount of water (86,400 litres a day) is the energy; the flow at any one moment is the power (1 litre). If you need more than the available flow, your power demand exceeds the power supply and cannot be met. If you dam and store the flow, then suddenly release it, you vastly increase the power; but you cannot increase the total energy.

In my view, the primary constraint is the total usable energy flow. The rate at which energy can be transferred (power) acts as a secondary constraint, particularly for renewables, because of their low energy density and intermittent availability. Conversely, a high power density does not necessarily translate into a high energy density. To illustrate: lightning has an electric power equivalent to the output of ten large nuclear power stations. But because it lasts for just a fraction of a second, it only provides enough energy to power an electric car for less than 1 kilometre.

The book also creates confusion because many of Smil’s examples of power density are actually energy densities. He is aware that power density as a measure should not focus on brief bursts of extreme energy availability, such as the solar electricity generated at noon in southern Arizona; it should be averaged over a larger territory and longer periods. And he appropriately calculates average, representative power densities, expressed as watts per square metre ( $\text{W m}^{-2}$ ). The standard unit for power, the watt, is defined as joules per second. But Smil’s power densities are averages calculated over an entire year, and so become energy densities (in watt-years per year), an equivalence he fails to explain. However, he is not alone in this confusing use of power units that actually denote energy flows. Climate scientists also use  $\text{W m}^{-2}$  to describe the radiative balance of the planet and its alteration by increasing concentrations of greenhouse gases.

There are also important qualitative aspects that need to be

➔ **NATURE.COM**

For more on science  
in culture, see:  
[nature.com/  
booksandarts](http://nature.com/booksandarts)



considered. It matters whether power refers to electricity or biomass. And it matters whether land is used exclusively or only partly for energy. Smil raises this issue but does not address it systematically. He could have used the concept of exergy (energy's ability to perform useful work) to differentiate between high-quality energy (such as electricity, which is versatile) and low-quality energy (such as straw, which demands costly conversions to become usable beyond burning). However, his comparisons look only at power densities, irrespective of quality. Smil sketches out a valuable taxonomy of energy-related land uses along two dimensions. The first is exclusivity: the site of a power plant, for instance, is unusable for agriculture; right-of-way land underneath transmission lines is not. The second is longevity of use: a nuclear-waste repository will be in place for centuries, whereas an annual crop such as maize (corn) for conversion to ethanol can be grown in rotation. Yet

**"Cities will require vast renewable-energy hinterlands."**

Smil does not use this taxonomy, and as a result sometimes compares 'apples and oranges'. Unlike in some of Smil's other books, the production quality of *Power Density* is regrettably low. I prefer readable graphics accompanied consistently by source and data referencing. In this book, many graph labels are hardly legible, and figures with references are the exception. Sometimes sources are mentioned in the text but not in the caption; at others, they are not even in the text. Graphs plot data on population and energy use, but statistical data sources are not specified or referenced.

*Power Density*'s detailed examination of the spatial constraints of energy options adds to Smil's earlier, pioneering treatment of the subject, making it useful for energy specialists interested in exploring a massive ramp-up of renewables. But its technical nature and language make it rather inaccessible to a wider audience. And its failure to explain fundamental concepts such as the difference between power and energy, and to provide adequate data and source referencing, make it unsuitable as a textbook. ■

**Arnulf Grubler** is at the International Institute for Applied Systems Analysis in Laxenburg, Austria, and the School of Forestry and Environmental Studies at Yale University in New Haven, Connecticut.  
e-mail: gruebler@iiasa.ac.at

## HUMAN EVOLUTION

# The cradle of humankind revisited

**Michael Cherry** catches up with new developments and old dilemmas at South Africa's hominin-fossil hotspot.

**I** reach the Cradle of Humankind after half an hour's drive from Johannesburg, through the Gauteng Highveld of South Africa. This open, grassy space scattered with trees is a World Heritage Site, riddled with limestone caves and hominin fossils.

A substantial chunk of the evidence that Africa is the wellspring of humanity was discovered here; and with anthropologist Lee Berger of the University of Witwatersrand (Wits) in Johannesburg set to unearth more at the Malapa and Rising Star Cave sites, the Cradle still rocks.

As a heritage site in a developing country, the area is a focus for national pride. But developments there are spurring questions over which part of the nation they serve.

Palaeontologists will rejoice over the launch, on 21 July, of a state-of-the-art vault to house star local finds, an adjunct to Wits's Centre of Excellence for Palaeosciences. The vault will allow specimens to be compared with other finds, both hominin and non-hominin, from around Africa. These include the Taung skull (*Australopithecus africanus*), dated to between 2 million and 3 million years ago, which was discovered north of Kimberley in 1924; specimens of *Australopithecus sediba* discovered at Malapa, including a remarkably complete skeleton called MH1, as well as casts of East African discoveries such as Lucy (*Australopithecus afarensis*) and the type specimen of *Homo habilis*, found by anthropologists Mary and Louis Leakey. The vault's laboratory has a micro-CT scanner and 3D printing facilities. But it is strictly for researchers' use.

What is there for the public? The Maropeng visitor centre opened a decade ago as an interpretation centre for the Sterkfontein Caves, site of the discovery of the 'Mrs Ples' fossil (*Australopithecus africanus*) in 1947 and, 50 years later, Little Foot, the most complete early-hominin skeleton known, which is as-yet undescribed. And late last year, a light, moveable, steel structure known as the Beetle was placed over the Malapa site to let the paying public



*Australopithecus sediba* was discovered at Malapa Cave in South Africa.

view excavations, once they resume at the site. (Digging has been on hold since 2009, when the remains of four *A. sediba* individuals were removed.)

The Beetle protects the delicate limestone system from rain, and lets wild animals move freely below. Standing on eight clavicle-like supports, it has a fabric roof that collects rainwater and channels it to a sanitation system. Visitors will watch excavations from a raised circular walkway. A pulley below the platform is attached to a hoist capable of bearing a tonne of rock.

But there are questions over how 'public' the Beetle actually is. Costing half a million US dollars — paid for largely through the National Research Foundation (using taxpayers' money), as well as Wits and the Gauteng provincial government — the site is in a private game reserve and the tourists, when they come, will probably be rich. The Maropeng centre demonstrates this. Built at a cost of US\$29 million, it charges \$13 for admission (around half that for students), which prices out many in a country where one-fifth of the people still live on \$28 a month. That could be reflected in Maropeng's visitor numbers. Planned to accommodate 1 million visitors a year, it receives between 230,000 and 250,000 and runs at an annual loss, picked up by the provincial government. Kruger National Park, by contrast, charges different rates for South Africans and foreign tourists, and receives 1.4 million visitors annually.

Remarkable fossils continue to emerge in the Cradle, and it presents no less remarkable opportunities for palaeotourism. But a way must be found to make the specimens widely accessible. *In situ* interpretation of australopithecine remains should present a uniquely uplifting experience for all, rich and poor. ■

**Michael Cherry** is in the Department of Botany and Zoology at Stellenbosch University in South Africa.  
e-mail: mic@sun.ac.za

# Magic bullets to blockbusters

**Marian Turner** delves into a history of the rapid rise of monoclonal antibodies.

In 2003, freshly graduated from university, I made a monoclonal antibody. After weeks of waiting for the mice that I had immunized with *Acanthamoeba* parasites to mount a robust immune response, I fused some of their spleen cells to cells derived from a mouse myeloma cancer. The aim: to generate immortal hybridoma cells, from which I could purify an endless stream of specific antibodies that bound to the parasites.

The protocol I used was a barely refined version of one described by biologists César Milstein and Gorges Köhler in the United Kingdom almost three decades before. Hybridoma technology spawned the broad-reaching field of monoclonal antibodies, which is celebrated by historian of medicine Lara Marks in *The Lock and Key of Medicine*. Essential reagents in any cell-biology laboratory, monoclonal antibodies are now also common in medicine, from pregnancy testing to blood typing and disease diagnostics.

Marks begins by summarizing scientists' early attempts to understand protective immunity and vaccination. The term magic bullets, coined by German physician Paul Ehrlich (who died 100 years ago next month) in his 1897 description of antibodies, is held up as an early beacon of researchers' hopes for antibodies in medicine. As Marks shows, the hypotheses of pioneering scientists, amazingly, have often only just missed the mark. Ehrlich's description of a cell producing 'side chains' that break off as antibodies in response to encountering foreign substances is remarkably close to our current knowledge of surface immunoglobulins and secreted antibodies. The book's early chapters also chronicle how physicians of the 1920s and 1930s developed serum-based therapies decades before anyone understood the basic mechanisms of antibody action.

There were high hopes for the revenue potential of monoclonal antibodies, but not everyone was convinced at first. Marks describes how the National Research and Development Corporation in the United Kingdom initially failed to see the practical applications. As a result, the original hybridoma technology was not patented. In 1979 and 1980, US scientists won patents for essentially equivalent technology using myeloma cells provided by Milstein — a development that left many in Britain chagrined, including then-prime minister Margaret Thatcher.

Such early hiccups were soon replaced



The fundamental protocol for making monoclonal antibodies has changed little in 40 years.

by focused commercialization, leading to an exponential rise in patents and several intellectual-property battles. Although some scientists have found this unpalatable, antibody-related royalties have been ploughed back into medical research — by 2012, the UK Medical Research Council's patents alone had earned £486 million (US\$770 million).

The possible ways of modifying antibodies have surged in the 40 years since Milstein and Köhler published their hybridoma protocol. Marks discusses the generation of antibody fragments, chimaeric antibodies and 'humanized' antibodies — products from non-human cells that have been modified to reduce unwanted immune responses. The unfolding antibody story paralleled advances in recombinant-DNA technology and transgenic animal models, which made such alterations possible. Yet it is mind-boggling that the first humanized antibodies were made before the advent of DNA-amplification technology.



**The Lock and Key of Medicine: Monoclonal Antibodies and the Transformation of Healthcare**  
LARA V. MARKS  
Yale Univ. Press: 2015.

'Humanized' also describes what Marks has done for the antibody story. *The Lock and Key of Medicine* presents rich details of who at which institute collaborated with whom on which scientific advance or commercialization process, and when. The meticulous accounts sometimes blur, but they convey a deep sense of the cumulative thought and effort embodied in today's antibody technologies, and remind us that interdisciplinary and international collaborations are not new. Readers in the field will appreciate the attention paid to defining episodes, such as the HLDA workshops starting in 1982, which resulted in classification and verification systems that brought coherence to the expanding catalogue of monoclonal antibodies.

In the closing chapters, Marks describes some of the monoclonals that have become blockbuster drugs: rituximab, infliximab and trastuzumab. The stories of these antibodies reflect how serendipitous clinical outcomes, such as cancer drugs successfully treating arthritis, have led to deeper understanding of the biology of both classes of disease.

There are some surprising gaps. There is no mention of the catastrophic 2006 phase I clinical trial of the monoclonal antibody TGN1412, manufactured by TeGenero to treat cancer and autoimmune diseases. The drug induced multiple organ failure in six healthy volunteers, caused by an unanticipated extreme immune reaction called a cytokine storm. The episode led to a revision of European guidelines for first-in-human trials. Also missing is a discussion of monoclonal antibodies that block the immune regulatory proteins PD-1 or CTLA-4, which are arguably the hottest up-and-coming agents in cancer therapy today, or of the use of HIV-specific monoclonal antibodies for treatment and vaccine design.

But it is a fast-paced field. The history of monoclonal antibodies ricochets between basic science, the clinic and the commercial world, and *The Lock and Key of Medicine* documents how lessons from one sphere have repeatedly led to advances in another. Marks concludes by pointing out that monoclonal antibodies have received less fanfare than their biotechnological peers, genetic engineering and stem cells. Her thorough telling of this rich history goes some way towards restoring the balance. ■

**Marian Turner** is a senior News & Views editor for Nature.



# Correspondence

## Taxonomic glory easier on eBay?

One of zoology's highest honours may now, it seems, be purchased on eBay (see [go.nature.com/ziq152](http://go.nature.com/ziq152)). For a few thousand dollars, you are offered the privilege of naming a 'small, rare' species. A species name will last forever, says the vendor — even as taxonomists themselves struggle to survive.

Taxonomists invest months confirming that a specimen is new to science. They sift through obscure literature — often in a different language and lamentably illustrated. More months are spent on the species' description, which must be accurate enough to enable future taxonomists (should they survive the sixth mass extinction) to confirm that their 'new' species is different. Eventually, they publish their work in a systematics journal with an impact factor typically below 2 — even when the species is a previously undescribed mammal (the olinguito *Bassaricyon neblina*, pictured; see K. M. Helgen *et al. ZooKeys* 324, 1–83; 2013).

These low impact factors make it hard for taxonomists to land positions in academia, and job opportunities in museums are sparse. Selling perpetuity on eBay is starting to look like an attractive alternative.

**Giovanni Strona** *European Commission, Joint Research Centre, Institute for Environment and Sustainability, Ispra, Italy.*  
[giovanni.strona@jrc.ec.europa.eu](mailto:giovanni.strona@jrc.ec.europa.eu)

## Europe needs Ebola outbreak consortium

The European Commission (EC) has been criticized for failing to define specific research pathways for tackling the recent outbreak of Ebola virus in West Africa (J. M. Martin-Moreno *et al. Lancet* 384, 1259; 2014). In our view, three changes would improve research into new interventions.

First, the EC needs to set up



a cooperative framework for implementing research activity during outbreaks. Experience indicates that existing networks have suboptimal capability to involve African institutions and local authorities in the research that could help to contain the epidemic (S. Lanini *et al. Lancet Infect. Dis.* 15, 738–745; 2015).

Second, the EC must put in place resources and infrastructure so that, in the event of an unexpected resurgence of the virus, intervention studies can be rapidly approved.

Third, research institutions need to cooperate more closely with one another. Research networks have already been set up (see [go.nature.com/9xtemn](http://go.nature.com/9xtemn)) and others have successfully supported fragile local health-care services. These include the European Mobile Laboratory Project, Quality Assurance Exercises and Networking on the Detection of Highly Infectious Pathogens, and the network of biosafety level 4 laboratories, Euronet P4.

We suggest that an inclusive and committed European Consortium should be established. This should carry out research between epidemics and promptly translate the results into actions during epidemics. In our view, this consortium would be most effective if it were self-sufficient. Crucially, its budget structure

would allow easy access to funds for implementing interventions. **Alimuddin Zumla** *University College London, London, UK.* **David Heymann** *Chatham House Centre on Global Health Security, London, UK.*

**Giuseppe Ippolito** *National Institute for Infectious Diseases, Rome, Italy.*  
[giuseppe.ippolito@inmi.it](mailto:giuseppe.ippolito@inmi.it)

## Animal studies must be useful, says public

The European Commission (EC) responded last month to 'Stop Vivisection', a European Citizens' Initiative to phase out animal testing, which was signed by more than one million people. The EC confirmed that it will not replace the existing directive on the protection of animals used for scientific purposes (2010/63/EU), which already matches the level of protection in countries with the most demanding legislation. Yet the citizens' principal argument relates not to animal suffering, but to the limited usefulness of results from animal models.

Citizens concerned about animal welfare may still accept research that is perceived as being of ultimate benefit to humans, but only if it delivers relevant results — a view that evidently helped to mobilize signatures in this case. This is one of only three citizens' initiatives since 2012 that

have gathered enough signatures to reach the EC, so the scientific community needs to take its criticisms seriously.

As long as blinding, randomization and appropriate sample sizes are not standard practices in animal research, claims of maximizing its benefits are not credible. Unreliable data from poorly designed studies and publication bias give misleading results on the therapeutic value of candidate drugs, leading to disappointing clinical trials (see, for example, S. Perrin *Nature* 507, 423–425; 2014). Researchers must aim to do research that stands up to critical scrutiny from all quarters.

**I. Anna S. Olsson**, **Nuno H. Franco** *Institute for Molecular and Cell Biology, Porto, Portugal.*  
[olsson@ibmc.up.pt](mailto:olsson@ibmc.up.pt)

## A prescient view of women in evolution

The remarkable nineteenth-century German biologist August Weismann (*Nature* 522, 31–32; 2015) also took a prescient stand in the discourse on the role of women in evolution.

Weismann challenged a popular theory of heredity proposed by US zoologist William K. Brooks in *The Law of Heredity* (Murphy, 1883). On the basis of the Lamarckian idea of an inheritance of acquired characteristics, Brooks argued that the 'hereditary force', or *Vererbungskraft* (Weismann's translation), was stronger in men than in women, writing that "something within the animal compels the male to lead and the female to follow in the evolution of new breeds". Weismann roundly refuted this idea, pointing out that children inherit as many characteristics from their mothers as from their fathers (A. Weismann *Die Bedeutung der Sexuellen Fortpflanzung für die Selektions-Theorie*; Fischer, 1886). **U. Kutschera** *Institute of Biology, University of Kassel, Germany.*  
[kut@uni-kassel.de](mailto:kut@uni-kassel.de)

## PALAEONTOLOGY

# Hallucigenia's head

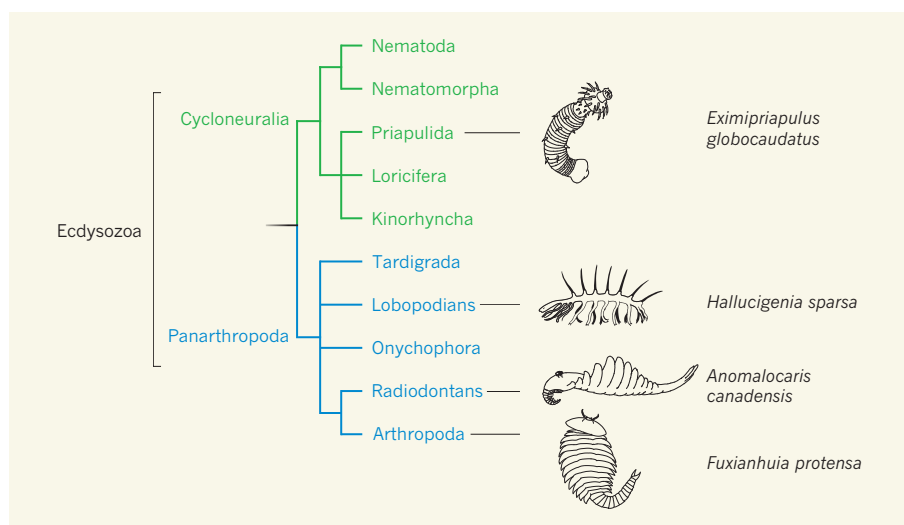
The finding of pharyngeal teeth and circumoral mouthparts in fossils of the Cambrian lobopodian animal *Hallucigenia sparsa* improves our understanding of the deep evolutionary links between moulting animals. **SEE LETTER P.75**

XIAOYA MA

Fossils provide direct evidence of evolutionary history, and their unique morphological combinations can reveal crucial evolutionary links between extant taxa<sup>1</sup>. Most major animal phyla first appear in the fossil record during the Cambrian period, 541 million to 485 million years ago, and this early flowering of animal life has been termed the Cambrian explosion. Therefore, Cambrian fossils are particularly important for understanding the origin and early evolution of major animal groups. In this issue, Smith and Caron<sup>2</sup> (page 75) redescribe one of the most celebrated Cambrian animals, *Hallucigenia sparsa*, and document several new features of this species, including its pharyngeal teeth and circumoral elements, which are suggested to be two of the few morphological characters uniting all groups within the Ecdysozoa.

The Ecdysozoa is far and away the richest animal group<sup>3</sup>. It is composed of eight extant phyla that shed their cuticle periodically to accommodate growth<sup>4</sup> — nematode worms and crustaceans are familiar examples. The two commonly recognized subgroupings of ecdysozoans, Cycloneuralia and Panarthropoda, have distinctly different body plans (Fig. 1). Cycloneuralia unites worm-like organisms (Nematoda, Nematomorpha, Priapulida, Kinorhyncha and Loricifera) that have a non-segmented body terminating in a mouth that can turn inside out (eversible) and has a ring of nerves behind it — their brain. By contrast, panarthropods (Arthropoda, Onychophora and Tardigrada) are all segmented, with paired legs, and have a dorsal (upper side) brain in front of the mouth. These great morphological disparities have made it difficult to illuminate the last common ancestor of the Ecdysozoa and to fully understand the evolutionary relationships between its phyla, particularly between Cycloneuralia and Panarthropoda. Early ecdysozoan fossils are crucial for addressing these questions.

Among the earliest Cambrian fossils, ecdysozoans are the most diverse and abundant group. They are best shown in exceptionally preserved Cambrian fossil localities, such as the Chengjiang biota in China (around 518 million years old) and the Burgess Shale in Canada (about



**Figure 1 | The Ecdysozoa.** This large animal group comprises eight extant phyla and two informal extinct groups, the lobopodians and the radiodontans. According to their body plans, ecdysozoans are commonly recognized as two distinct subgroups, Cycloneuralia and Panarthropoda. The fossil record of both can be traced to the earliest Cambrian period, and some members, such as priapulids and arthropods, have changed little over 500 million years of evolution. Cambrian lobopodians and radiodontans represent crucial evolutionary links between Cycloneuralia and Panarthropoda. Smith and Caron<sup>2</sup> present an analysis of the lobopodian *Hallucigenia sparsa* that includes details of its head structures. (Illustrations by David Baines.)

508 million years old). The body plans of some of the organisms represented have not changed much over 500 million years of evolution, such as priapulids (commonly known as penis worms) and arthropods (jointed-legged invertebrates with an exoskeleton and a segmented body, such as insects and spiders). Other Cambrian groups are extinct but represent crucial evolutionary stages, such as lobopodians (an informal group of worm-like animals with non-jointed legs) and radiodontans (a group of animals characterized by possessing a pair of frontal appendages at the anterior part of the head and a ventral (lower side) mouth surrounded by radial tooth plates).

Cambrian lobopodians are assigned to Panarthropoda on the basis of their segmented body and paired legs, but they also share a worm-shaped soft body and a terminal mouth with cycloneurians. These unusual character combinations make Cambrian lobopodians particularly relevant for understanding the evolutionary links between the two major ecdysozoan groups.

*Hallucigenia sparsa* from the Burgess Shale is certainly the most famous Cambrian lobopodian animal. It was originally reconstructed upside down<sup>5</sup> and considered to be one of the most bizarre Cambrian creatures until it was recognized as a lobopodian animal armed with dorsal spines<sup>6</sup>. However, owing to lack of evidence of clear head structures, the front and rear ends of *H. sparsa* have been a subject of debate. Smith and Caron's redescription includes a new set of anatomical features that once and for all clarifies the anterior–posterior orientation of *H. sparsa*. They show that the animal had an elongated head with a pair of dorsal eyes. It also had hardened, lamellae-like structures surrounding its mouth opening (circumoral elements), and the front part of its foregut (its pharynx) was lined with teeth.

Although pharyngeal teeth and circumoral mouthparts have been reported in other Cambrian lobopodians<sup>7,8</sup>, Smith and Caron have provided the most convincing evidence yet of equivalent structures in this extinct group. These findings amplify the transitional



status of Cambrian lobopodians, because the pharyngeal teeth of *H. sparsa* most closely resemble those of Cambrian priapulids, whereas circumoral structures are also a key characteristic of Cambrian radiodontans. More crucially, *H. sparsa* is now regarded<sup>2,9</sup> as an ancestor of living onychophorans (commonly known as velvet worms), so the finding of *H. sparsa* mouthparts suggests that the absence of circumoral elements and pharyngeal teeth in extant onychophorans is probably the result of secondary loss. Thus, this combined structure is now reported for all major ecdysozoan groups.

Smith and Caron further notice the similarities of ecdysozoan mouthparts (see Supplementary Note 1, transformation series 9 and 13 of the paper<sup>2</sup>), and suggest that all pharyngeal teeth and circumoral structures across ecdysozoan groups share a single origin from the last common ancestor of ecdysozoans. This provides new anatomical features to unite the Cycloneuralia and Panarthropoda.

However, this conclusion is bound to provoke some controversy. Limited by the vagaries of preservation, it is difficult to determine the detailed morphology and symmetry of the pharyngeal teeth and circumoral elements of *H. sparsa* — such details are essential for further comparative studies. Although the homology of ecdysozoan pharynxes lined with teeth is well accepted, the evolutionary links between the circumoral structures of Cycloneuralia and Panarthropoda are less clear, because these differ substantially in their structure, relative position, construction and symmetry<sup>10</sup>. Therefore, a more complete understanding of the evolutionary origin and transformation sequence of these mouthparts depends on a more thorough comparison of their morphology, development and innervation across all ecdysozoan groups. For this, new fossil evidence showing transitional features of the mouthparts between cycloneuralians and panarthropods would be particularly enlightening. ■

**Xiaoya Ma** is at Yunnan University, Kunming 650091, China, and at the Natural History Museum, London, UK.  
e-mail: x.ma@nhm.ac.uk

1. Edgecombe, G. D. & Legg, A. D. in *Arthropod Biology and Evolution* (eds Minelli, A. et al.) 393–415 (Springer, 2013).
2. Smith, M. R. & Caron, J.-B. *Nature* **523**, 75–78 (2015).
3. Telford, M. J., Bourlat, S. J., Economou, A., Papillon, D. & Rota-Stabelli, O. *Philos. Trans. R. Soc. London Ser. B* **363**, 1529–1537 (2008).
4. Aguilardo, A. M. et al. *Nature* **387**, 489–493 (1997).
5. Conway-Morris, S. *Spec. Pap. Palaeontol.* **20**, 1–95 (1977).
6. Ramsköld, L. & Hou, X. *Nature* **351**, 225–228 (1991).
7. Hou, X. G., Ma, X. Y., Zhao, J. & Bergström, J. *Lethaia* **37**, 235–244 (2004).
8. Vannier, J., Liu, J., Lerosey-Aubril, R., Vinther, J. & Daley, A. C. *Nature Commun.* **5**, 3641 (2014).
9. Smith, M. R. & Ortega-Hernández, J. *Nature* **514**, 363–366 (2014).
10. Dewel, R. A. & Eibye-Jacobsen, J. *Hydrobiologia* **558**, 41–51 (2006).

This article was published online on 24 June 2015.

## NANOTECHNOLOGY

# Colourful particles for spectrometry

A smartphone camera, patterned with arrays of filters made from colloidal suspensions of coloured particles, has been transformed into a powerful tool for spectral analysis. [SEE LETTER P.67](#)

**NORM C. ANHEIER**

In 1857, Michael Faraday gave a well-attended lecture at the Royal Institution of Great Britain, during which he presented his pioneering experimental work on the interaction of light with matter<sup>1</sup>. Faraday's study probed the fundamental properties of light related to its reflection and absorption by progressively smaller particles. During the presentation, very fine gold particles dispersed in liquid were shown to produce vivid colours not seen with larger particles. Faraday did not know that he had created suspensions of particles now known as colloidal quantum dots (CQDs), but, guided by insight, he concluded that the distinct colours were due to the minute sizes of the gold grains. On page 67 of this issue, Bao and Bawendi<sup>2</sup> describe how they have exploited the unique optical properties of CQDs to develop a compact optical spectrometer that could be integrated with a smartphone camera or used as a miniature, hand-held sensing tool.

Faraday had glimpsed a special condition that allows a particle's quantum nature to be expressed. His work set the course for

nanoscience and quantum theory, but it took 125 years before the physics of the phenomenon that he observed was attributed to quantum size effects<sup>3</sup>. It is now known that, when CQDs are exposed to light, some of the electrons in these particles are excited as they gain energy from the photons. However, unlike large particles and bulk materials, the nanoscale dimensions of the quantum-dot

particles confine the electrons and change the energy difference between their excited and relaxed states. CQDs emit light when the electrons relax from a higher to a lower energy state (Fig. 1). The colour of the light depends on the states' energy difference and is critically linked to the size of the particles, which can be controlled when producing the CQDs. The physics that underpins this behaviour allows CQDs to be used for spectroscopy.

The first simple spectrometer, consisting of a dispersive prism, was developed by Isaac Newton, who proved that white light is composed of a spectrum of many colours<sup>4</sup>. These days, optical spectrometers have become indispensable instruments used to measure the distribution of light's colours (wavelengths) in a variety of complex scientific investigations. Astronomers use them to collect and analyse optical spectra of exoplanets that may have life-supporting atmospheres<sup>5</sup>. Planetary scientists are using spectrometers on board rovers on the



**Figure 1 | Colloidal quantum dots.** When they are excited by ultraviolet light (pictured), colloidal suspensions of minuscule particles (known as colloidal quantum dots, or CQDs) fluoresce at different colours depending on the particle size. Bao and Bawendi<sup>2</sup> have exploited the unique optical absorptive properties of CQDs to develop a compact spectrometer that serves as a powerful tool to analyse the spectral characteristics of light.

surface of Mars to analyse the composition of soil and rocks, looking for clues to the planet's past environment and whether conditions may have been favourable for microbial life<sup>6</sup>. Optical spectrometers routinely support activities that underpin our daily lives, such as biomedical research, drug discovery, renewable energy, forensic science, environmental monitoring and chemical detection.

The optical spectrometers used in these applications tend to be complex and costly because of their numerous high-precision optical and mechanical components and the stringent requirements for the alignment of these parts. They can also suffer from poor throughput, because much of the input light is scattered or absorbed as it passes through many components before reaching the detector for analysis. Finally, to be able to distinguish, or resolve, two nearby wavelengths, the instruments must typically be large.

Bao and Bawendi have overcome many of these limitations through an elegant integration of nanotechnology with the image sensors used in digital cameras. Their spectrometer is based on a design involving broadband absorptive filters, which are similar to the coatings applied to sunglasses to block ultraviolet light. The filters are made from a series of CQDs, each with a specific particle size.

This design concept can be explained by considering the case of a single sensor and a broadband absorptive optical filter. The sensor detects visible light and the filter has a specific, known (measured independently) cut-off wavelength, below which light is totally absorbed. Ideally, the filter should efficiently transmit light above this cut-off wavelength.

Now consider an additional sensor and an absorptive filter tuned to another, slightly different, cut-off wavelength. Both pairs of sensors and filters are illuminated with light of unknown colour content. The difference in the signals registered by the two sensors is a measure of the incident light power between the different cut-off wavelengths. In principle, extending this approach to greater numbers of sensors with filters tuned to different cut-off wavelengths will increase the range of colours that can be measured and the ability to resolve two adjacent colours.

Bao and Bawendi achieved this scaling by applying 195 different broadband absorptive CQD filters to hundreds of locations on the pixelated image sensor and by using an extended spectrum-reconstruction method to handle the large set of sensor readings.

Commercial nanotechnology developments have led to simplified quantum-dot synthesis and precise control of dot size. CQDs can be used as tailorable broadband absorptive filters, because both the spectral absorption and the emission properties vary with particle

size. It is now feasible to produce CQDs with continuously and finely varying absorption cut-off colours, from the deep violet to near-infrared wavelengths. These solutions can be directly patterned on the image-sensor pixels using inkjet or direct-contact printing methods. The long-term stability of the patterned quantum dots has also improved to provide reasonable device lifetime. It is these key factors that enabled Bao and Bawendi to develop their CQD spectrometer, and the simplicity of their design overcomes the constraints usually seen in conventional spectrometers.

Future developments in nanotechnology and their potential for commercialization may provide the full complement of CQD materials needed for spectral measurements beyond the visible range. One promising research area concerns chalcogenide CQDs, which have optical emission and absorption properties that extend to longer, infrared wavelengths<sup>7</sup>. Further technical challenges must be overcome to improve CQD materials and reduce optical losses. If practical automated quantum-dot

patterning on image sensors can be realized, then the costs that limit widespread integration of this technology into consumer electronics will be reduced. In the future, we may see tiny, high-resolution CQD spectrometers used on space missions or as ubiquitous sensing elements in household devices connected to the Internet. ■

**Norm C. Anheier** is at the Pacific Northwest National Laboratory, Richland, Washington 99352, USA.  
e-mail: norm.anheier@pnnl.gov

1. Faraday, M. *Phil. Trans. R. Soc. Lond.* **147**, 145–181 (1857).
2. Bao, J. & Bawendi, M. G. *Nature* **523**, 67–70 (2015).
3. Ekimov, A. I. & Onushchenko, A. A. *JETP Lett.* **34**, 345–349 (1981).
4. Newton, I. *Opticks* (Smith & Walford, 1704).
5. Kaltenecker, L. *et al. Astrobiology* **10**, 89–102 (2010).
6. Wiens, R. C., Maurice, S. & the ChemCam Team. *Geochem. News* **145**, 41–48 (2011).
7. Kershaw, S. V. *et al. Chem. Soc. Rev.* **42**, 3033–3087 (2013).

## PUBLIC HEALTH

# The case for pay to quit

**A randomized controlled trial of four financial-incentive programmes for smoking cessation finds that reward-based schemes lead to sustained abstinence, but low public acceptability of such schemes threatens their adoption.**

**THERESA M. MARTEAU  
& ELENI MANTZARI**

**T**obacco remains the most lethal legal product<sup>1</sup>, killing up to half of all users<sup>2</sup> and accounting for more than half of the difference in life expectancy between rich and poor members of society<sup>3</sup>. Higher pricing, which equates to a financial penalty, is thought to be the most successful intervention for reducing smoking. A 50% increase in inflation-adjusted prices is estimated to cut consumption by 20% in high-, middle- and low-income countries, with the largest impacts on the young and on poor people<sup>4</sup>. For those wanting to quit, behavioural interventions and pharmacotherapy can help, although few attempts result in sustained quitting<sup>5</sup>. Financial incentives are a relatively new addition to the repertoire of behavioural interventions. Writing in the *New England Journal of Medicine*, Halpern *et al.*<sup>6</sup> present a trial of four incentive schemes that adds to the growing evidence base (Fig. 1).

The trial is the first to compare four incentive schemes for smoking cessation against usual care. It involved 2,538 employees of a US company, their relatives and friends. Two of the schemes targeted individuals and two

targeted groups of six participants who were incentivized through the collective performance of the group. One individual and one group-based scheme provided rewards of US\$800 for smoking cessation, and the other two schemes required refundable \$150 deposits, together with a \$650 reward, for successful participants.

Acceptance of the reward-based schemes — assessed as the proportion of participants enrolling in the scheme to which they were assigned — was 90%, much higher than for the deposit schemes, at 13.7%. There were no differences in the acceptance of individual and group-based schemes. When comparing the entire cohort of participants, regardless of whether or not they enrolled on the scheme they were offered (with those not enrolling in the scheme assumed to have remained smokers), the quit rates in the intervention groups at 6 months ranged from 9.4% to 16% — all higher than the 6% quit rate in the usual-care group. Quitting was higher for the reward- than for the deposit-based schemes (15.7% versus 10.2%) and similar for individual and group-based schemes.

Among the small proportion of participants who accepted the deposit schemes, 52% had quit at 6 months, compared with 17% of those





**Figure 1 | Carrots work for smokers.** Halpern *et al.*<sup>6</sup> show that reward-based financial-incentive schemes are more commonly accepted by smokers than are deposit-based schemes, but that both approaches lead to higher rates of quitting than usual care.

on reward schemes. This outcome led Halpern *et al.* to conclude that use of deposits is better. However, the lack of interest in these schemes limits their usefulness, as does the absence of sustained effects. At 12 months, 6 months after the incentives were stopped, about 50% of quitters were smoking again. Only those who had been enrolled in reward-based schemes, who retained quit rates of 7.5% for the individual and 8.7% for the group schemes, showed an advantage over the 3.4% quit rate at 12 months achieved through usual care.

The problem of ‘gaming’ — faking being a smoker to qualify for enrolment on a scheme or being a non-smoker to remain on a scheme — is of particular concern when using financial incentives for smoking cessation. Although reports of quitting by participants in Halpern and colleagues’ trial were validated by checking cotinine levels (a component of tobacco) in their saliva, smoking at enrolment was confirmed in only a small minority of participants. An estimated 20% of those recruited were non-smokers, although this did not affect the study’s findings. By contrast, a study of pregnant smokers who were offered participation in a financial-incentive scheme revealed that none of the 239 enrolled, all of whom were biochemically tested at baseline, were non-smokers<sup>7</sup>. This raises questions about the contextual features of incentive schemes that can be used to minimize gaming.

Halpern and colleagues’ well-designed study, the largest known of its kind so far, makes a considerable contribution to the growing evidence about financial incentives

for smoking cessation. Two recent systematic reviews<sup>8,9</sup>, one of which<sup>8</sup> includes the results of this trial, suggest that financial-incentive schemes can be effective in achieving sustained quitting, particularly when substantial incentives are used and when offered to pregnant smokers. The effect of incentives is also doubled in more-deprived populations<sup>9</sup>, suggesting that this approach motivates greater change in people on low incomes, among whom rates of smoking are not only markedly higher but also resistant to reduction. This finding highlights the potential contribution of financial incentives for reducing the health inequalities related to smoking, including the significantly higher death rates from tobacco among poor people.

Were such schemes a pill or a behavioural intervention not involving money, they would doubtless be included in the range of interventions on offer to smokers wanting to quit. We are, however, “funny about money” and in particular about “money out of place”<sup>10</sup>. Given their equal effectiveness to other interventions, financial incentives are less acceptable to the public, and in turn potentially to health professionals and policy-makers, than other interventions for changing behaviours<sup>11</sup>.

Why is there not greater acceptability of a seemingly cost-effective intervention for reducing the leading cause of preventable premature death worldwide? The low acceptability of paying people to stop smoking (or to lose weight or take medication) reflects several concerns about ideas of fairness, including

‘coercing the vulnerable’, ‘rewarding the feckless’ and ‘not rewarding the responsible’. The first of these often arises in the context of incentivizing the disadvantaged, because it is assumed that such individuals are less free to resist monetary temptations. This compromises the use of incentives for addressing health inequalities. An example of the effect of the latter two concerns is a situation in the United States<sup>12</sup>, in which objections from non-smoking employees led to the rejection by management of a \$750 reward scheme for smoking cessation (for which there was strong evidence of effectiveness) and the adoption of a \$625 penalty deducted from smokers’ salaries — an intervention for which at the time there was no evidence of effectiveness.

The acceptability of financial incentives can change: it increases with information about effectiveness and varies with incentive type (acceptability is lower for cash incentives and higher for voucher schemes)<sup>13</sup>. Nonetheless, financial interventions to change behaviour — whether they involve price increases through taxation or reward schemes — are much less acceptable than other, often less effective, interventions<sup>14</sup>. Studies are needed that explore the scale and duration of financial-incentive schemes in different populations of smokers, and that aim to identify the characteristics of schemes producing the greatest changes. In parallel with this, research is needed to develop ways of increasing the acceptability of interventions that could significantly help to reduce the death toll from tobacco, particularly among poor individuals. ■

**Theresa M. Marteau and Eleni Mantzari**  
are in the Behaviour and Health Research  
Unit, University of Cambridge,  
Cambridge CB2 0SR, UK.  
e-mail: tm388@cam.ac.uk

1. www.vox.com/2014/5/19/5727712/the-three-deadliest-drugs-in-america-are-all-totally-legal
2. Peto, R. *et al.* *Br. Med. Bull.* **52**, 12–21 (1996).
3. Jarvis, M. J. & Wardle, J. in *Social Determinants of Health* (eds Marmot, M. & Wilkinson, R.) 240–255 (Oxford Univ. Press, 1999).
4. Jha, P. & Peto, R. *N. Engl. J. Med.* **370**, 60–68 (2014).
5. US Department of Health & Human Services. *The Health Consequences of Smoking — 50 Years of Progress: A Report of the Surgeon General, 2014* (HHS, 2014).
6. Halpern, S. D. *et al.* *N. Engl. J. Med.* **372**, 2108–2117 (2015).
7. Ierfino, D. *et al.* *Addiction* **110**, 680–688 (2015).
8. Cahill, K., Hartmann-Boyce, J. & Perera, R. *Incentives for Smoking Cessation* (Cochrane Library, 2015).
9. Mantzari, E. *et al.* *Prevent. Med.* **75**, 75–85 (2015).
10. Sandel, M. J. *What Money Can't Buy: The Moral Limits of Markets* (Farrar, Straus & Giroux, 2012).
11. Promberger, M., Brown, R. C. H., Ashcroft, R. E. & Marteau, T. M. *J. Med. Ethics* **37**, 682–687 (2011).
12. Volpp, K. G. & Galvin, R. J. *Am. Med. Assoc.* **311**, 909–910 (2014).
13. Promberger, M., Dolan, P. & Marteau, T. M. *Soc. Sci. Med.* **75**, 2509–2514 (2012).
14. Diepeveen, S., Ling, T., Suhrcke, M., Roland, M. & Marteau, T. M. *BMC Public Health* **13**, 756 (2013).

## PLANETARY SCIENCE

# Sink holes and dust jets on comet 67P

Analyses of images taken by the Rosetta spacecraft reveal the complex landscape of a comet in rich detail. Close-up views of the surface indicate that some dust jets are being emitted from active pits undergoing sublimation. [SEE LETTER P.63](#)

PAUL WEISSMAN

When do 18 holes not make for a pleasant afternoon playing golf? When the 18 holes are located on the surface of a comet speeding through the Solar System. On page 63 of this issue, Vincent *et al.*<sup>1</sup> describe the holes, also called pits, that comprise one of the many discoveries of the European Space Agency's Rosetta mission to comet 67P/Churyumov-Gerasimenko (67P). The Rosetta spacecraft went into orbit around 67P in August 2014, and the surprises have been coming fast since then. Vincent *et al.* propose a mechanism for the formation of the pits and identify them as one of the sources of active dust jets.

Comets are the most primitive bodies in the Solar System; they are the remnants of its formation process. Comets therefore retain a physical and chemical record of the conditions and materials in the solar nebula — the gas and dust cloud out of which the Sun and planets formed 4.56 billion years ago. Conveniently, comets have spent most of that

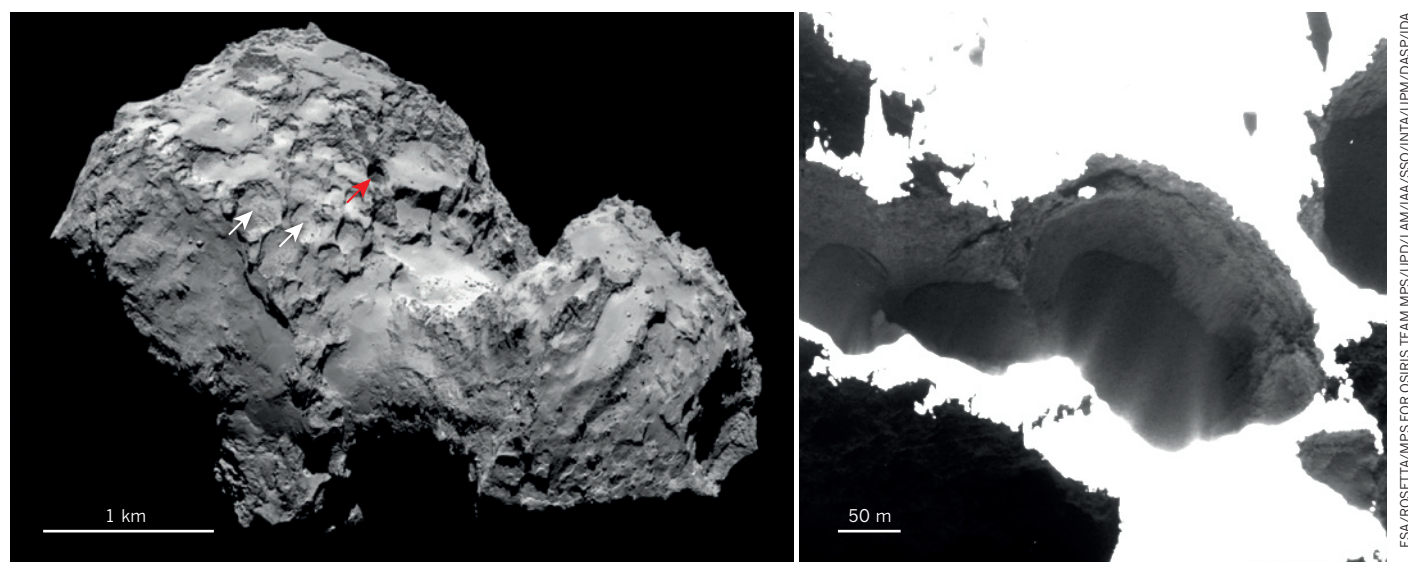
time in two very cold storage locations: the Kuiper belt beyond the orbit of Neptune and the spherical Oort cloud outside the planetary region, stretching halfway to the nearest stars. The distant Oort cloud is the source of the long-period comets that have orbital periods ranging up to millions of years. The Kuiper belt is the source of the Jupiter-family comets, such as 67P, which typically have periods of less than 20 years and orbital dynamics that are strongly affected by Jupiter.

As a comet approaches the Sun and warms up, the central solid part, known as the cometary nucleus (comprised of volatile ices and primitive meteoritic material), begins to sublimate and becomes enveloped by a freely outflowing atmosphere called the coma. One of the first surprises for Rosetta, the first ever comet-rendezvous mission, was the odd shape of the target comet's nucleus (Fig. 1a)<sup>2</sup>. Although some nuclei comprised of two large pieces and looking like a bowling pin had been observed before by fly-by missions to other comets, the two lobes of 67P sit on top of each other, with a narrow 'neck' in between.

There is intense speculation as to how this odd configuration may have formed. Did two cometary nuclei gently collide randomly in the solar nebula, or is the nucleus a single piece that has been oddly sculpted by sublimation processes? Although the former is the more likely scenario, some scientists on the mission suspect the latter.

Rosetta's camera system, the Optical, Spectroscopic and Infrared Remote Imaging System (OSIRIS), is comprised of narrow-angle and wide-angle digital cameras. As the OSIRIS team of scientists<sup>2</sup> began to map the surface of the nucleus using the cameras, they discovered 18 pits on the surface, which Vincent *et al.* now describe more thoroughly. The cometary nucleus has a diameter of approximately 4 kilometres. The pits are typically about 200 metres in diameter and about 180 metres deep. Pit-like features have been observed on other cometary nuclei, but the morphology of the pits on 67P has not been seen before. They typically have cylindrical shapes with circular openings and near-vertical walls (although at least one pit seems to be lying at a steep angle). And some of the pits are clearly active: images of pits that are illuminated by sunlight show dust jets emanating from their walls and/or floors (Fig. 1b).

How did the pits form? Vincent *et al.* suggest that they are 'sink holes', which formed when material near the surface of the nucleus collapsed into the low-density interior. Rosetta's Radio Science Investigation team has found<sup>2</sup> that the nucleus has an average bulk density of only  $470 \pm 45$  kilograms per cubic metre, about half the density of solid water ice. But the Grain Impact Analyser and Dust Accumulator instrument has measured<sup>3</sup> a dust-to-ice



**Figure 1 | The nucleus of comet 67P/Churyumov-Gerasimenko (67P).** Vincent *et al.*<sup>1</sup> analysed images of comet 67P taken by the Optical, Spectroscopic and Infrared Remote Imaging System cameras on the Rosetta spacecraft. **a**, The complex nucleus topography includes large, flat-floored basins (indicated by white arrows). A large, circular pit is visible just above the centre of the image (red arrow). **b**, A string of pits dot the surface of the

cometary nucleus. In active pits such as these, bright jets of dust are seen being emitted from the sunlit walls. The contrast of this image has been enhanced to highlight the interiors of the pits and the jets. As a result, the cometary surface looks very bright, but in reality it reflects only about 6% of the incoming sunlight — roughly the same as the black toner particles in a laser printer cartridge.



mass ratio of  $4 \pm 2$ , suggesting that silicates and organics, rather than ices, make up about 80% of the mass of the nucleus. This in turn implies that 75–85% of the nucleus interior is empty space, a parameter known as porosity. A high porosity is predicted by the leading scenarios for the internal structure of cometary nuclei, which suggest that they are aggregates<sup>4</sup> of smaller, icy bodies that gently came together in the solar nebula. These aggregates are also referred to as rubble piles<sup>5</sup>. This concept has provided insights into the behaviour of comets, such as random and other splitting events.

The morphology of 67P's surface is dominated in some areas by large, flat-floored basins, similar to features seen on the nucleus of comet<sup>6</sup> Wild 2. It has been suggested that these are sublimation basins that slowly widen as the walls sublime, leaving large, non-volatile particles that cover the basin floor. The basins cannot be impact craters because they have the wrong size distribution (there are too many large ones), and because not many impact craters are expected on a small cometary nucleus such as 67P.

Could the pits described by Vincent *et al.* be the precursors of the basins, slowly widening as their walls sublime? Many of the pits found by OSIRIS are located in the same region on the nucleus where many of the large sublimation basins are found. Both comet 67P and comet Wild 2 are relatively young — that is, they have only recently (within the past 60 years) been perturbed by the gravitational field of Jupiter to perihelion distances (the point in their orbit closest to the Sun) at which it is warm enough for water ice in the nucleus to sublime, and at which the activity that manifests itself as the bright cometary coma and tails begins. If this is so, why are sublimation basins not observed on other, perhaps older, Jupiter-family comets such as Tempel 1 and Hartley 2? Older nuclei may have accumulated thicker layers of non-volatile materials that have buried the sublimation basins and substantially lowered the activity levels of those comets.

Rosetta has already indicated that it has more surprises for us. On 13 June 2015, the orbiter began receiving signals from the Philae lander, which is on the surface of the comet nucleus and was last heard from in November 2014. With its batteries recharging, Philae probably has much more information to transmit about its final landing location. Also, the activity of the nucleus is expected to reach a maximum soon after the comet passes through perihelion at 1.25 astronomical units from the Sun (a point about 25% farther from the Sun than Earth's orbit) in mid-August 2015. Rosetta will then follow 67P away from the Sun as cometary activity begins to wane. What changes will we see on the nucleus surface? And how will this alien golf course look from Rosetta's vantage point then? ■

Paul Weissman is at the Jet Propulsion Laboratory, NASA, Pasadena, California 91109, USA.

e-mail: weissman@jpl.nasa.gov

1. Vincent, J.-B. *et al.* *Nature* **523**, 63–66 (2015).
2. Sierks, H. *et al.* *Science* **347**, aaa1044 (2015).

3. Rotundi, A. *et al.* *Science* **347**, aaa3905 (2015).
4. Donn, B. & Hughes, D. in *20th ESLAB Symp. Exploration of Halley's Comet* (eds Battrick, B. *et al.*) 523–524 (ESA, 1986).
5. Weissman, P. R. *Nature* **320**, 242–244 (1986).
6. Kirk, R. *et al.* in *46th Lunar and Planetary Science Conf. Abstr.* 2244 (2015).

## EVOLUTION

# Reptile sex determination goes wild

Wild populations of an Australian lizard have sex chromosomes and also exhibit temperature-controlled sexual development, providing insight into how these two sex-determining mechanisms may evolve back and forth. [SEE LETTER P.79](#)

JAMES J. BULL

Vertebrate sex determination is getting interesting. On page 79 of this issue, Holleley *et al.*<sup>1</sup> report elaborate field and laboratory studies on an egg-laying Australian lizard, *Pogona vitticeps*, and reveal that its sex is determined both by its complement of chromosomes and by the temperature at which its eggs are incubated. Earlier reports<sup>2,3</sup> had hinted at the possibility of combined genotypic and environmental sex determination in some lizards, but it had never before been convincingly reported in the wild.

Fifty years ago, the dominant view, laid out by Susumo Ohno<sup>4</sup>, was that there was an inexorable evolutionary progression from genetically labile mechanisms of sex determination to highly refined and differentiated sex chromosomes. Ohno suggested that

this progression was recapitulated across the vertebrate groups, with sex chromosomes becoming increasingly more entrenched as one moved up what was then perceived as the evolutionary phylogenetic ladder — fish had labile systems, mammals and birds had entrenched sex-chromosome systems, and reptiles were in the middle of this transition.

Within two decades, that view had radically changed<sup>5</sup>. By then, many reptiles were known to have full-blown sex-chromosome systems, whereas many others had sex determination by incubation temperature and no hint of inherited sex. It was also realized that the evolution of sex determination follows basic evolutionary principles, and that chromosomal and environmental sex determination can both be highly functional, adaptive systems. In other words, they are not different steps along an



**Figure 1 | *Pogona vitticeps*.** Holleley *et al.*<sup>1</sup> show that, in this lizard species, embryos that have two Z chromosomes and are thus genetically male can develop as female at warm egg-incubation temperatures.

evolutionary progression, but are alternative states that could, in theory, evolve back and forth. At the time, however, such transitions were unknown, and it seemed that each reptile species had one type of system or the other, like two sides of a coin.

This latter perception is now also changing, and it seems that we are seeing the coin as it flips. The lizards studied by Holleley *et al.* (Fig. 1) have visibly recognizable sex chromosomes with female heterogamety — females have a Z and a W chromosome, and males have two Zs. However, the authors find that nearly 20% of ZZ individuals sampled in the wild are female instead of male. Incubation of eggs in the laboratory revealed that the ZZ offspring develop as male at low temperatures but that an increasing proportion develop as female as the incubation temperature increases. So ZZ females in the wild probably come from warm nests.

These observations cement previous speculation about how sex chromosomes and environmental sex determination may coexist and how the transition between them may occur. At first glance, two problems are created by a system that combines sex chromosomes and environmental sex determination. The first is that arbitrary environmental determination of sex would lead to ZW males and ZW females, which when mated would sometimes lead to WW offspring. If the W is a degenerate sex chromosome, meaning it has lost many of its functional genes, WW offspring would be inviable or sterile. This is solved in *P. vitticeps* by the simple fact that ZW is always female — only the ZZ genotype becomes either sex — so there is no possibility of a ZW–ZW mating and thus no WW genotype.

The second problem is that the consistent development of some ZZ individuals into females creates an excess of females in the population. This dilemma is solved through sex-ratio selection, which automatically adjusts the frequency of the W chromosome to progressively lower levels as more ZZ females are produced. Provided that the environment is neither too warm nor too cool, the equilibrium population may sit indefinitely at a point that includes some ZW females and some ZZ females. But there is a continuum of equilibria spanning from pure chromosomal sex determination to pure environmental sex determination as the average nest temperature increases, and it is a steep transition. Extended Data Figure 4 of Holleley and colleagues' paper<sup>1</sup> shows that the wild population sits on a virtual cliff of the changeover between the two mechanisms.

The study augments this picture with several other observations. First, the temperatures causing ZZ lizards to become female are slightly lower for offspring of ZZ mothers than for offspring of ZW mothers. This result suggests underlying quantitative variation in the propensity for environmental sex

determination, a satisfying confirmation of theory: those individuals most genetically disposed to develop as ZZ females have offspring that are also genetically disposed to become ZZ female. Second, ZZ females have markedly higher fecundity than ZW females. This result was not expected, and although it is not contradictory to theory, it raises the question of why. The answer may shed light on the selective advantage of temperature-dependent sex determination.

Holleley and colleagues' findings will no doubt inspire parallel work on other species, especially in efforts to understand the transitions between sex-determining mechanisms and to explore the ecological and evolutionary consequences of the different mechanisms. The ability to assess the fitness of ZZ and ZW females raised at the same temperature will enable comparisons to be made that are crucial to understanding the relative advantages of the two systems and the possible costs of sex-chromosome degeneration.

Broader geographic and longitudinal comparisons for these lizards will give insight into the ramifications of climate change on this temperature-dependent reproductive mode. However, the established equilibrium between genetically and environmentally determined sex in these lizards should respond quickly to climate change, because an overproduction of ZZ females in warm years would lead to a compensatory reduction in the frequency of ZW

females in the next generation and beyond.

The findings in this one system should dovetail with recent revelations of frequent changes in lizard heterogametic determination, a discovery made possible by easy genome sequencing of less-studied species<sup>6</sup>. The accumulating information about the molecular bases of reptile sex determination<sup>7</sup> will add greatly to this understanding, and may reveal interesting constraints imposed on the transitions<sup>8</sup>. The emerging picture is that reptilian sex determination is more flexible on an evolutionary scale than could ever have been imagined. ■

**James J. Bull** is at the Institute for Cellular and Molecular Biology, the Center for Computational Biology and Bioinformatics and the Department of Integrative Biology, University of Texas at Austin, Austin, Texas 78712, USA.

e-mail: bull@utexas.edu

- Holleley, C. E. *et al.* *Nature* **523**, 79–82 (2015).
- Radder, R. S., Quinn, A. E., Georges, A., Sarre, S. D. & Shine, R. *Biol. Lett.* **4**, 176–178 (2008).
- Quinn, A. E. *et al.* *Mol. Genet. Genomics* **281**, 665–672 (2009).
- Ohno, S. *Sex Chromosomes and Sex-linked Genes* (Springer, 1967).
- Janzen, F. J. & Paukstis, G. L. *Q. Rev. Biol.* **66**, 149–179 (1991).
- Gamble, T. *et al.* *Mol. Biol. Evol.* **32**, 1296–1309 (2015).
- Janes, D. E. *et al.* *Biol. Lett.* **10**, 20140809 (2014).
- Schwanz, L. E., Ezaz, T., Gruber, B. & Georges, A. *J. Evol. Biol.* **26**, 2544–2557 (2013).

## NEUROBIOLOGY

# Inversion in the worm

**Combinations of spatially and temporally restricted transcription factors are shown to coordinate movement in nematode worms by controlling the formation of synaptic connections to and from motor neurons. SEE LETTER P.83**

**VILAIWAN M. FERNANDES  
& CLAUDE DESPLAN**

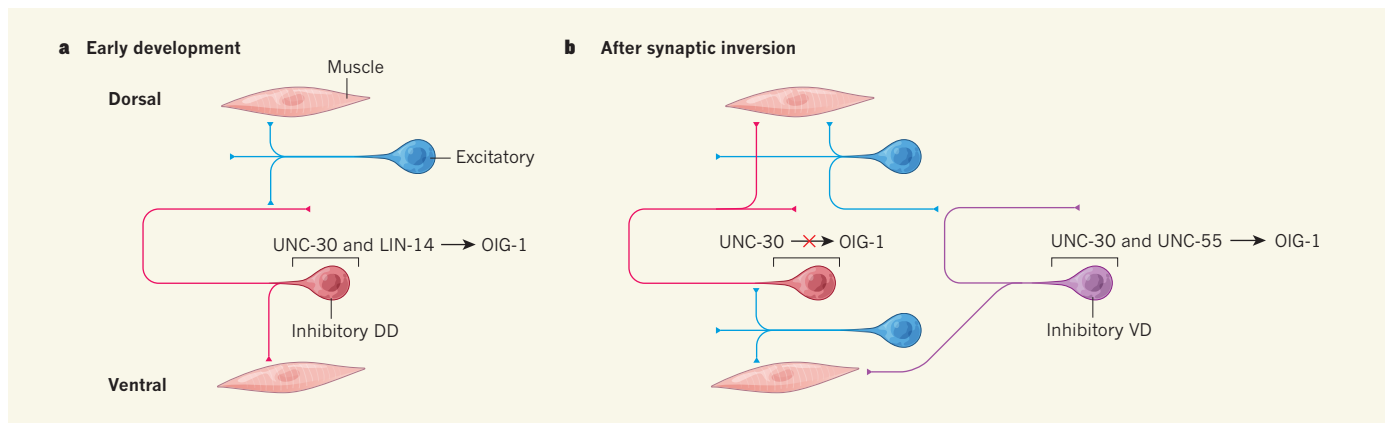
**N**ervous systems are staggeringly complex. To generate appropriate behavioural outputs, the countless synaptic connections that neurons form with other cells must be precisely regulated to ensure that they are arranged in the right circuits at the right time. On page 83 of this issue, Howell *et al.*<sup>1</sup> show that, in the simple neuronal circuits of the nematode worm *Caenorhabditis elegans*, such spatio-temporal precision is achieved through transcription factors that function together to restrict the expression of a newly identified synaptic organizer protein called OIG-1.

Nematode movement involves coordinated muscle contractions that are regulated by complex interactions between motor neurons on

the worm's dorsal (upper) and ventral (lower) sides. Two classes of inhibitory D-type motor neuron<sup>2</sup> in particular help to control this process. Dorsal D (DD) neurons make synaptic connections to muscles on the dorsal side of the worm, and are themselves innervated by excitatory motor neurons called cholinergic neurons from the animal's ventral side. By contrast, ventral D (VD) neurons innervate ventral muscles and receive synaptic inputs from dorsal cholinergic neurons. Activation of ventral cholinergic neurons therefore contracts ventral muscles and activates DD neurons, inhibiting dorsal-muscle contraction, whereas activation of dorsal cholinergic neurons leads to contraction of dorsal, but not ventral, muscles.

DD neurons form during embryonic development, but VD neurons arise after a larva's exoskeleton has moulted for the first time<sup>2</sup>.





**Figure 1 | Transcriptional regulation of synaptic specificity.** **a**, During early development of nematodes, excitatory motor neurons (blue) on the dorsal side of the embryo make synaptic connections with, and so excite, dorsal muscle and inhibitory dorsal D-type (DD) motor neurons (red). DD neurons innervate, and thus inhibit contraction of, ventral muscle. Howell *et al.*<sup>1</sup> report that the synaptic inputs and outputs of early DD neurons are controlled by LIN-14. This transcription factor, together with the ubiquitously expressed transcription factor UNC-30, promotes expression of the protein OIG-1,

which prevents the formation of synaptic outputs from DD neurons to dorsal muscle cells. **b**, After the larva's first moult, LIN-14 is no longer expressed and DD neurons undergo a synaptic inversion — they become innervated by ventral excitatory neurons and themselves innervate dorsal muscle. Ventral D-type (VD) motor neurons (purple) innervate ventral muscle cells and express the transcription factor UNC-55, which, together with UNC-30, promotes OIG-1 expression and prevents VD neurons from forming inhibitory connections to dorsal muscle.

In fact, before VD neurons develop, early DD neurons innervate ventral muscles and receive synaptic inputs from motor neurons on the dorsal side of the animal<sup>2</sup> (Fig. 1a). They later undergo a synaptic inversion when VD neurons arise (Fig. 1b). Howell *et al.* investigated the regulatory logic behind this synaptic rewiring using worms that harbour mutations in the gene *unc-30*, which encodes an evolutionarily conserved transcription factor, UNC-30. This protein is expressed in all D-type motor neurons throughout development, and is regarded as a 'terminal-identity selector' — its expression ultimately defines whether neurons will become D-type motor neurons<sup>3–5</sup>.

In addition to the uncoordinated locomotion for which the gene is named, the authors found that *unc-30* mutant worms had defects in synaptic specificity. Although the cell bodies (the nucleus-containing regions) of D-type motor neurons are appropriately positioned in *unc-30* mutants, the early DD and VD neurons fail to innervate ventral muscles and instead make abnormal synaptic connections to dorsal muscles. Moreover, VD neurons do not receive synaptic inputs from dorsal cholinergic motor neurons.

How does UNC-30 regulate synapse formation in specific spatio-temporal patterns, given its ubiquitous expression in D-type neurons throughout development? Previous studies<sup>6–9</sup> have shown that mutations in two other genes encoding transcription factors, *lin-14* and *unc-55*, recapitulate different aspects of the defects found in *unc-30* mutants. Expression of LIN-14 is temporally restricted to early development, and mutation of *lin-14* leads to the abnormal formation of dorsal synapses from early DD neurons<sup>8,9</sup>. By contrast, UNC-55 is restricted to VD neurons, where it regulates synaptic specificity to ventral muscles<sup>6,7</sup>.

Howell *et al.* reasoned that UNC-30 might

cooperate with LIN-14 and UNC-55 to promote the expression of a molecule that blocks dorsal synaptic output. This molecule would be expected to be expressed under the control of UNC-30 acting with LIN-14 in early DD neurons, and under the control of UNC-30 acting with UNC-55 in VD neurons, inhibiting the formation of synaptic outputs on the dorsal side of the animal. But under this simple model, ventral synaptic outputs would remain normal in *unc-30* mutants. Because this is not the case, additional layers of regulation must be involved.

OIG-1, a member of the immunoglobulin superfamily, whose members mediate interactions between cells, is a putative target of UNC-30, LIN-14 and UNC-55 (ref. 10). Howell and colleagues provide evidence that *oig-1* is expressed in early DD neurons and is later restricted to VD neurons. Furthermore, its expression is altered by perturbations to *lin-14* and *unc-55* expression. The authors report that loss of *oig-1* leads to the formation of dorsal synaptic outputs similar to those seen in *unc-30*, *lin-14* and *unc-55* mutants. Moreover, synaptic innervation of early DD and VD neurons by cholinergic neurons on the dorsal side of the animal is disrupted in *oig-1* mutants. This suggests that OIG-1 coordinates both the inputs to and outputs of D-type motor neurons.

Howell *et al.* found that OIG-1 is located along the ventral side of early DD and VD neurons. Given that the protein can organize dorsal synaptic inputs and outputs, this observation suggests that it acts indirectly. The authors also show that forced expression of *oig-1* could not block synaptic rewiring of late DD motor neurons to ventral muscles, indicating that other factors must cooperate with OIG-1 to regulate synaptic specificity. Identifying these cofactors, which, like *oig-1*,

must be differentially expressed in early DD and VD neurons, will be of great interest, as will determining the molecular mechanisms by which they act with OIG-1 to regulate synaptic inputs and outputs through pre- and postsynaptic partner molecules.

Achieving appropriate synaptic specificity involves many developmental steps that act together to ensure that neurons assume the correct identity. Determinants of neuronal identity, which are regulated by terminal-identity selectors, include the production of particular neurotransmitter molecules, the guidance of nerve fibres in specific directions and the appropriate growth of branched projections called dendrites. Howell and colleagues' work demonstrates that the formation and targeting of synapses are also traits that give neurons a particular identity, and that synapses, too, can be regulated by terminal-identity selectors such as UNC-30. ■

**Vilaiwan M. Fernandes and Claude Desplan** are in the Department of Biology, New York University, New York, New York 10003, USA. e-mail: cd38@nyu.edu

- Howell, K., White, J. G. & Hobert, O. *Nature* **523**, 83–87 (2015).
- White, J. G., Albertson, D. G. & Anness, M. A. R. *Nature* **271**, 764–766 (1978).
- Eastman, C., Horvitz, H. R. & Jin, Y. J. *Neurosci.* **19**, 6225–6234 (1999).
- Jin, Y., Hoskins, R. & Horvitz, H. R. *Nature* **372**, 780–783 (1994).
- Cinar, H., Keles, S. & Jin, Y. *Curr. Biol.* **15**, 340–346 (2005).
- Walthall, W. W. & Plunkett, J. A. J. *Neurosci.* **15**, 1035–1043 (1995).
- Hedgecock, E. M., Culotti, J. G., Hall, D. H. & Stern, B. D. *Development* **100**, 365–382 (1987).
- Ruvkun, G. & Giusto, J. *Nature* **338**, 313–319 (1989).
- Hallam, S. J. & Jin, Y. *Nature* **395**, 78–82 (1998).
- Aurelio, O., Hall, D. H. & Hobert, O. *Science* **295**, 686–690 (2002).

# The architecture of the spliceosomal U4/U6.U5 tri-snRNP

Thi Hoang Duong Nguyen<sup>1</sup>, Wojciech P. Galej<sup>1</sup>, Xiao-chen Bai<sup>1</sup>, Christos G. Savva<sup>1</sup>, Andrew J. Newman<sup>1</sup>, Sjors H. W. Scheres<sup>1</sup> & Kiyoshi Nagai<sup>1</sup>

**U4/U6.U5 tri-snRNP is a 1.5-megadalton pre-assembled spliceosomal complex comprising U5 small nuclear RNA (snRNA), extensively base-paired U4/U6 snRNAs and more than 30 proteins, including the key components Prp8, Brr2 and Snu114. The tri-snRNP combines with a precursor messenger RNA substrate bound to U1 and U2 small nuclear ribonucleoprotein particles (snRNPs), and transforms into a catalytically active spliceosome after extensive compositional and conformational changes triggered by unwinding of the U4 and U6 (U4/U6) snRNAs. Here we use cryo-electron microscopy single-particle reconstruction of *Saccharomyces cerevisiae* tri-snRNP at 5.9 Å resolution to reveal the essentially complete organization of its RNA and protein components. The single-stranded region of U4 snRNA between its 3' stem-loop and the U4/U6 snRNA stem I is loaded into the Brr2 helicase active site ready for unwinding. Snu114 and the amino-terminal domain of Prp8 position U5 snRNA to insert its loop I, which aligns the exons for splicing, into the Prp8 active site cavity. The structure provides crucial insights into the activation process and the active site of the spliceosome.**

The protein-coding sequences of most eukaryotic genes are interrupted by non-coding segments called introns. Introns are removed from precursor mRNAs (pre-mRNAs) and the flanking coding segments (exons) are spliced together to form mRNAs by two successive trans-esterification reactions within a dynamic multi-megadalton protein-RNA complex known as the spliceosome. This complex comprises five canonical subunits, namely U1, U2, U4, U5 and U6 snRNPs, and numerous non-snRNP factors<sup>1</sup>. Each snRNP contains an snRNA, seven Sm or LSm proteins, and a number of snRNP-specific proteins. During the initial stages of spliceosome assembly, U1 and U2 snRNPs recognize the pre-mRNA 5' splice site and branch point, forming pre-spliceosomal A complex. The subsequent binding of the pre-assembled U4/U6.U5 tri-snRNP allows formation of the fully assembled spliceosomal B complex, which is converted to the catalytically active B\* complex through extensive structural and compositional remodelling. During this process, U4 and U6 snRNAs, which are extensively base-paired in tri-snRNP, are unwound, U1 and U4 snRNPs are released, and many new proteins join the spliceosome<sup>1–3</sup>. This leads to the formation of a highly structured RNA network between U2, U5 and U6 snRNAs and the 5' splice site and branch point sequences in the pre-mRNA. The extensively base-paired U2–U6 snRNAs harbour catalytic magnesium ions<sup>4</sup> and position the branch point and 5' splice site for the first trans-esterification reaction, which produces exon 1 and lariat intron–exon 2 intermediates. Further remodelling to C complex enables U5 snRNA loop I to align exons 1 and 2 for nucleophilic attack of exon 1 at the 3' splice site, yielding spliced mRNA and lariat intron products<sup>5,6</sup>. Finally, the spliceosome is disassembled before the next round of splicing.

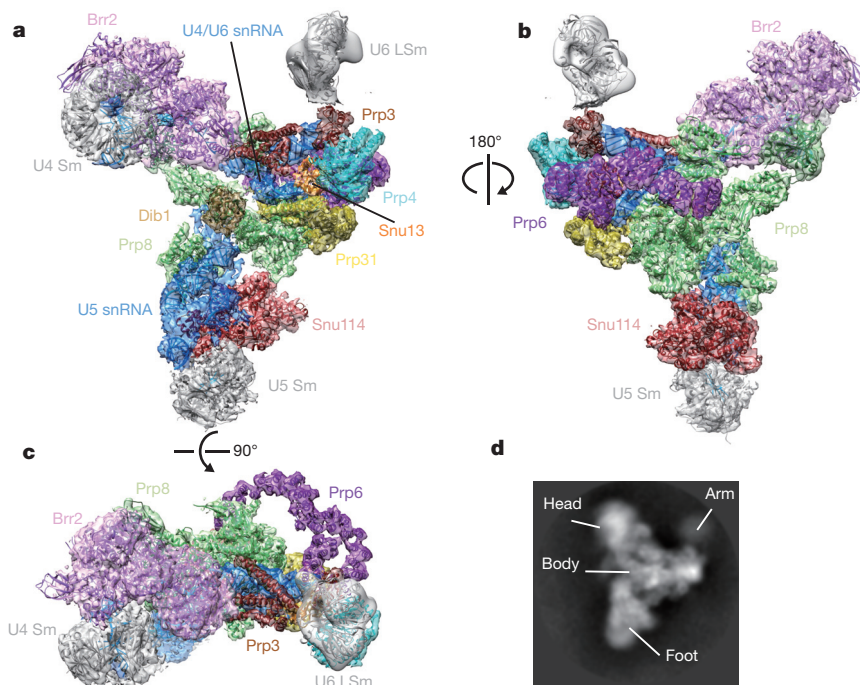
U4/U6.U5 tri-snRNP is the largest pre-assembled spliceosomal complex, containing U5 snRNA, extensively base-paired U4/U6 snRNAs and over 30 proteins<sup>7,8</sup> (Extended Data Table 1). Three key proteins, Prp8, Brr2 and Snu114, have crucial roles in activation of the spliceosome and formation of the active site<sup>1</sup>. Prp8 forms crosslinks

with 4-thiouridine introduced at key positions within U5 and U6 snRNAs and the substrate pre-mRNA, showing that Prp8 is involved in substrate positioning and closely associated with the catalytic RNA core<sup>9,10</sup>. Brr2 helicase, the activity of which is regulated by the GTPase Snu114 (refs 11–13), catalyses the unwinding of the U4/U6 snRNA duplex<sup>14,15</sup>. Interactions between tri-snRNP proteins have been investigated by yeast two-hybrid and *in vitro* binding assays<sup>16,17</sup>. Electron cryo-microscopy (cryoEM) reconstruction of crosslinked human tri-snRNP at 21 Å resolution revealed a tetrahedral overall shape with no clear domain separation<sup>18</sup>. Negative stain microscopy of cross-linked yeast tri-snRNP revealed a triangular shape with maximum dimension of 30–34 nm (ref. 19). The highly biased orientation of tri-snRNP on carbon films precluded full three-dimensional analysis while the projection structure revealed three extruding domains termed head, foot and arm; the arm domain adopts variable positions with respect to the rest. Some key proteins were localized within the projection structure using genetically introduced tags<sup>19</sup>. Brr2 and U4/U6 snRNP were attributed to the head and arm domains, respectively. On the basis of this, it was proposed that Brr2 may engage with U4/U6 snRNAs for unwinding when Snu114—mapped in the hinge region—brings the arm and head domains closer<sup>19</sup>.

The development of high-speed direct electron detectors<sup>20,21</sup> and powerful maximum likelihood algorithms for classification and particle alignment<sup>22</sup> have made it possible to determine the structure of macromolecular assemblies at near-atomic resolution by cryoEM<sup>23</sup>. By applying these new methods we obtained a map of native unstained yeast tri-snRNP at an overall resolution of 5.9 Å in which protein  $\alpha$ -helices and RNA double helices are readily discernible. This enabled us to fit the double-stranded helices of U5 snRNA and U4/U6 snRNAs as well as previously determined crystal structures or homology models of nearly all the proteins. The structure accounts for a wealth of biochemical and genetic data from yeast and human spliceosomes, and suggests a possible mechanism for B complex formation and the activation of the spliceosome.

<sup>1</sup>MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge CB2 0QH, UK.





**Figure 1 | Overview of the U4/U6.U5 tri-snRNP structure with its protein and RNA components modelled into cryo-EM density.** **a**, Front view, facing concave surface; **b**, back view; **c**, top view. **d**, 2D class average showing the different domains of tri-snRNP: head, body, arm and foot.

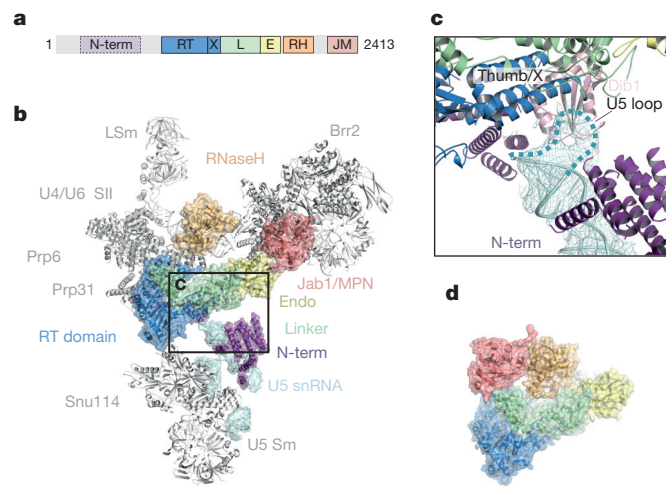
### CryoEM of the tri-snRNP complex

U4/U6.U5 tri-snRNP was purified from yeast by a gentle procedure without crosslinking, and the sample was subjected to cryoEM analyses (Methods and Extended Data Fig. 1). Using a combination of statistical classification and movie processing<sup>22,24</sup> (Extended Data Fig. 2), we obtained a density map with an overall resolution of 5.9 Å by the ‘gold standard’ Fourier shell correlation (FSC) = 0.143 criterion<sup>25</sup> with local resolution ranging from 5.0 Å to 20 Å (Extended Data Fig. 3 and 4; Methods). The map revealed clear densities for double-stranded RNA, with protein helices appearing as long tubes and  $\beta$ -sheets as flat densities (Supplementary Video 1). The density for the LSm proteins in the flexible arm domain became clearer after using a new multi-body refinement method (Methods and Extended Data Fig. 3).

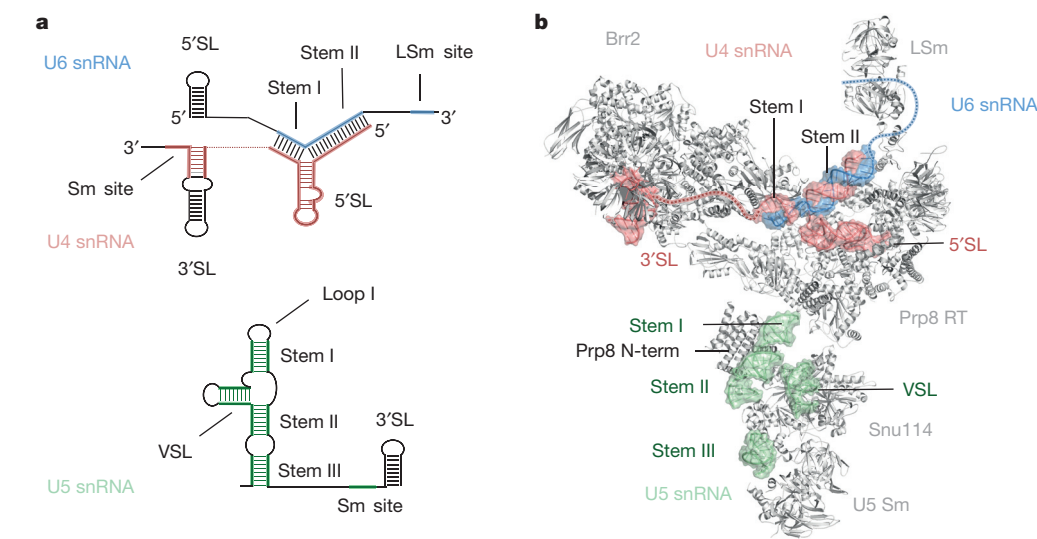
### Overall structure

Yeast U4/U6.U5 tri-snRNP has an overall Y-shape with a maximum dimension of approximately 300 Å (Fig. 1). The large domain of Prp8 (residues 885–1,824), consisting of the reverse transcriptase-like (RT), linker and type II endonuclease-like domains<sup>10</sup>, is located near the centre of the assembly and its crystal structure was fitted into the map as a rigid body<sup>10</sup> (Fig. 2 and Extended Data Fig. 5a). The orientation of the RNaseH-like domain with respect to the large domain is inverted in tri-snRNP as compared with the Prp8–Aar2 complex<sup>10</sup> (Fig. 2d). Three segments of clear double-stranded RNA density extending from Prp8 to the foot domain are assigned to co-axially stacked stems I and II, variable stem–loop and stem III of U5 snRNA (Fig. 3; Extended Data Fig. 6) connected to the U5 Sm core (Extended Data Fig. 5e). Snu114 shows a significant sequence similarity to eukaryotic translation elongation factor 2 (EF2)<sup>11,26</sup> comprising domains I–V (Fig. 4 and Extended Data Figs 5c and 7). Homology models of each domain of Snu114 (residues 120–1,008) were fitted individually into the density adjacent to Prp8 and U5 snRNA, revealing a contact between domain III of Snu114 and the RT domain of Prp8 (Fig. 4). The structure of the N-terminal 884 residues of Prp8 is still unknown. The N-terminal helix of the RT domain of Prp8 (RT $\alpha$ 1)<sup>10</sup> extends further in tri-snRNP towards a bundle of four long helices. Another cluster of long helices, which makes close contact with the co-axially

stacked stems I and II of U5 snRNA, is found in the vicinity (Fig. 3b and Extended Data Fig. 6c). The region containing residues 420–542 of Prp8 is known to interact with the N-terminal half of Snu114 (ref. 27), and 4-thiouridine introduced at C79 of U5 snRNA crosslinks with both Prp8 and Snu114 (ref. 28), suggesting that the density adjacent to Snu114 and U5 snRNA is part of the Prp8 N terminus. At the tip of stem I the density assigned to the U5 loop I extends towards the RT thumb/X domain of Prp8 and makes close contact with a thioredoxin-like fold of Dib1 (Fig. 2c and Extended Data Fig. 5f)

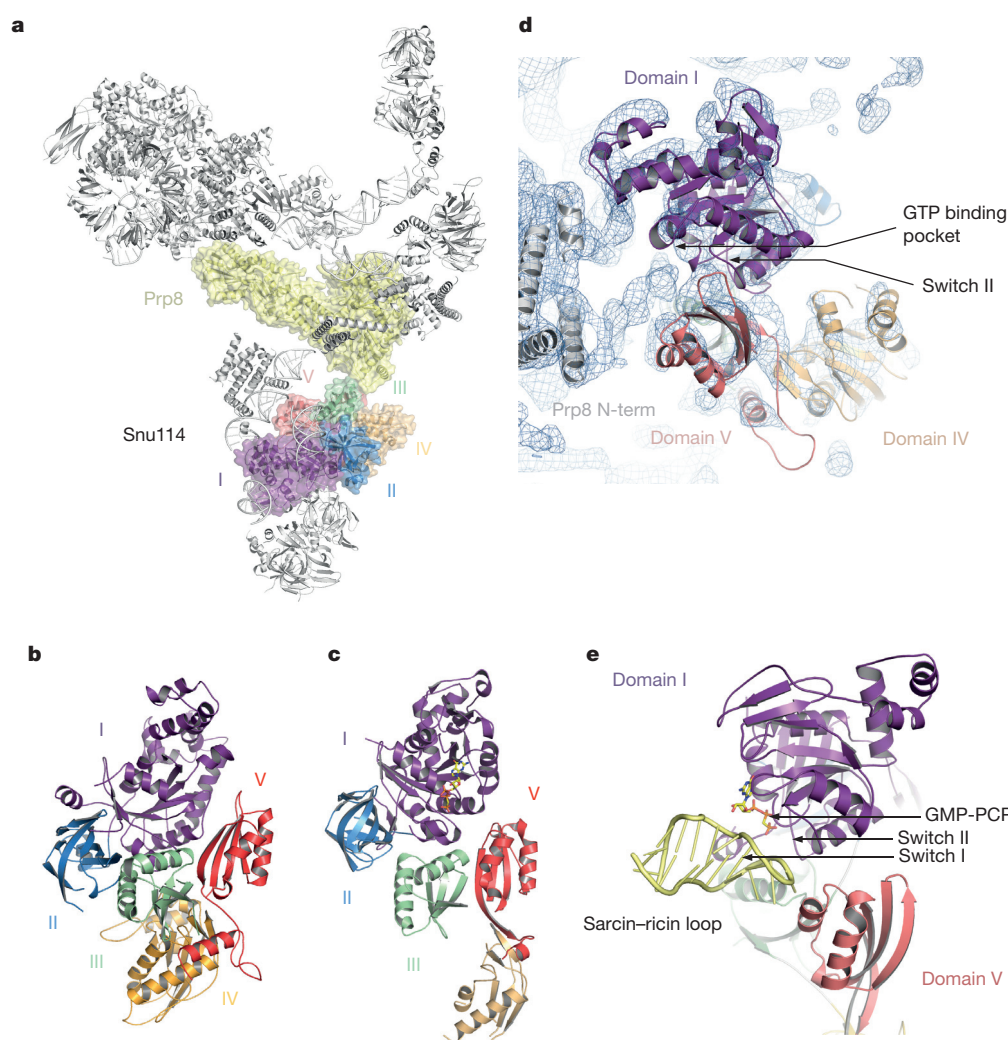


**Figure 2 | Prp8 in tri-snRNP.** **a**, Domain organization of Prp8. The structure of the N-terminal domain (residues 1–884) is unknown. RT, reverse transcriptase-like domain; X, thumb/X; L, linker; E, endonuclease-like domain; RH, RNaseH-like domain; JM, Jab1/MPN domain. **b**, The large domain of Prp8 is located at the centre of tri-snRNP. The Jab1/MPN domain is bound to Brr2 (refs 31, 32). **c**, Loop I of U5 snRNA is inserted into the active site cavity and in contact with Dib1. **d**, Prp8 in the Prp8–Aar2 complex<sup>10</sup> is shown with its large domain in the same orientation as in **b**. In tri-snRNP, the RNaseH domain is inverted while the Jab1/MPN domain in complex with Brr2 is located at the opposite end of the large domain.



**Figure 3 | The snRNA components of U4/U6.U5 tri-snRNP.**

**a**, Secondary structures of U4/U6 and U5 snRNAs. **b**, Double-stranded regions of U4/U6 and U5 snRNAs modelled into the cryo-EM map. SL, stem-loop.



**Figure 4 | Structure of Snu114 in tri-snRNP.**

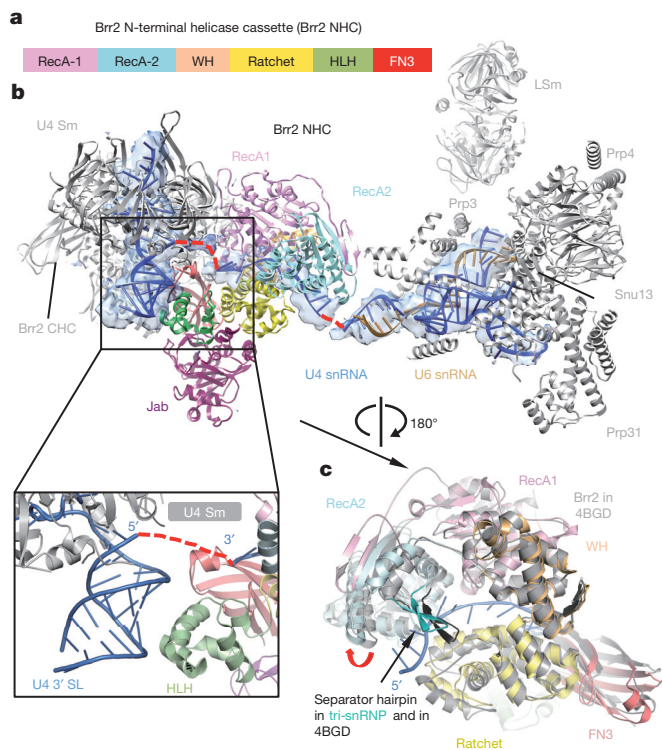
**a**, Location of Snu114 in the U4/U6.U5 tri-snRNP. **b**, Arrangement of domains (I–V) in Snu114 (see Extended Data Fig. 7). **c**, Domain arrangement in EF-G bound to the ribosome<sup>44</sup>. **d**, The interface between the N-terminal domain of Prp8 and Snu114. Some of the uninterpreted density at the interface may be attributed to the unmodelled switch I loop. **e**, The interaction of the switch region of EF-G with the sarcin-ricin loop<sup>44</sup> for GTPase activation.

(ref. 29). This is in good agreement with the fact that a 16-kDa protein is crosslinked to 4-thiouridine incorporated at U97 in the U5 snRNA loop I (ref. 28). The binding of Dib1 is further stabilized by the N-terminal helices of Prp8 (Fig. 2c).

Brr2 forms a stable complex with the Jab1/MPN domain of Prp8 (refs 30–32), and its characteristic shape was recognized in the

less-ordered head domain (Figs 1 and 5 and Extended Data Fig. 5b). Although this part of the map is lower in resolution, the individual domains of yeast Brr2 were fitted into the density together with the Jab1/MPN domain<sup>31</sup>. This revealed a widening of the gap between the two RecA domains of the N-terminal cassette (Fig. 5c). Co-axially stacked stems I and II of U4/U6 snRNAs and the 5' stem-loop of





**Figure 5 | Brr2 mode of unwinding.** **a**, Domain organization of Brr2 N-terminal helicase cassette (NHC). WH, winged helix; HLH, helix-loop-helix; FN3, fibronectin3-like domains. The inactive C-terminal helicase cassette (CHC) has the same domain organization. **b**, U4/U6 di-snRNP and its interaction with Brr2 in tri-snRNP. The domains of Brr2 NHC are coloured as in **a**. The single-stranded RNA between U4/U6 stem I and U4 snRNA 3' stem-loop is already loaded in the active site of Brr2. When the Hel308 structure<sup>43</sup> is overlaid onto the NHC of Brr2, its 10-nucleotide DNA substrate coincides with the density in the Brr2 active site, which extends to U4 snRNA 3' stem-loop (red dotted line). The helix-loop-helix domain of Brr2 interacts with U4 snRNA 3' stem-loop (inset). **c**, Superposition of the RecA1 domain of Brr2 in the crystal structure<sup>31</sup> (Protein Data Bank (PDB) accession 4BGD, in grey) and in tri-snRNP (domains coloured as in **a**) shows the opening of the gap between the RecA1 and RecA2 domains (indicated by the red arrow) to accommodate the RNA substrate.

U4 snRNA branching from the three-way junction are unambiguously identified near the N-terminal helicase cassette of Brr2 and Prp8 (Figs 1 and 3 and Extended Data Fig. 6). Snu13 and the Nop domain of Prp31 bind to the kink-turn at the tip of the U4 snRNA 5' stem-loop<sup>33</sup>. Initially the crystal structures of Snu13 and Prp31 were fitted individually around the kink-turn but the density clearly showed that the coiled-coil domain of Prp31 is rotated by approximately 60° with respect to the Nop domain in tri-snRNP (Fig. 1 and Extended Data Fig. 5g). The four-helix bundle of the coiled-coil domain is in contact with the RT domain of Prp8 (Fig. 1). Furthermore, Prp31 used for crystallization was a truncated form<sup>33</sup> and in our map clear  $\alpha$ -helical density extends from both N and C termini (Extended Data Fig. 5g). The C-terminal helix extends from the 5' stem-loop to the three-way junction, in agreement with previous foot-printing experiments<sup>34</sup> (Fig. 1a). Prp4 contains seven WD repeats at the C terminus<sup>35</sup> and its characteristic seven-bladed  $\beta$ -propeller domain is packed against Snu13 and the junction between the Nop and coiled-coil domain of Prp31 (Fig. 1 and 5 and Extended Data Fig. 5j). Four additional  $\alpha$ -helices, probably belonging to the N terminus of Prp4, were built into the rod-like density on top of the  $\beta$ -propeller. Prp3 is predicted to have a ferredoxin-like domain at its C terminus<sup>36</sup> and interacts with Prp4 and U4/U6 stem II (ref. 37).

Density sandwiched between the Prp4 WD40 domain and U4/U6 stem II, two long helices lying along U4/U6 stem II and a number of connected helices nearby probably belong to Prp3 (Fig. 1 and Extended Data Fig. 5i). The U4 core domain is wedged between the tandem helicase cassettes of Brr2 (Figs 1a, c and 5b and Extended Data Fig. 6a). The 3' stem-loop of U4 snRNA contacts the helix-loop-helix domain of the N-terminal helicase cassette (Fig. 5b), which contains several lysine/arginine residues close to the RNA backbone of the 3' stem-loop of U4 snRNA. On the basis of the previous labelling data<sup>19</sup>, U6 LSm proteins are fitted into the flexible arm region in the multi-body refined map (Fig. 1a and Extended Data Figs 3c and 5l).

A striking elongated curved  $\alpha$ -solenoid density bridging the RNaseH-like domain of Prp8 and the WD40 domain of Prp4 is assigned to the tetratricopeptide (TPR) motifs of Prp6 (Fig. 1b, c and Extended Data Fig. 5d). Prp6 is required for the accumulation of tri-snRNP<sup>38</sup> and is proposed to act as a bridge between U5 and U4/U6 snRNPs<sup>16,39</sup>. Prp6 contains up to 19 predicted TPR motifs, each comprising a helix-loop-helix motif<sup>39</sup> and 37 connected idealized poly-alanine helices were built into the map. Nine canonical tandem TPR motifs at the C terminus of the protein form a highly curved  $\alpha$ -helical solenoid-like structure, which contacts Snu13, U4 snRNA 5' stem-loop and the Prp4 WD40 domain in tri-snRNP (Fig. 1b, c). This is consistent with the fact that antibodies against the C-terminal fragment of human Prp6 immunoprecipitate U5 snRNP but not tri-snRNP, as the C-terminal domain in our structure is in close contact with U4/U6 snRNP, which presumably occludes the epitope<sup>39</sup>.

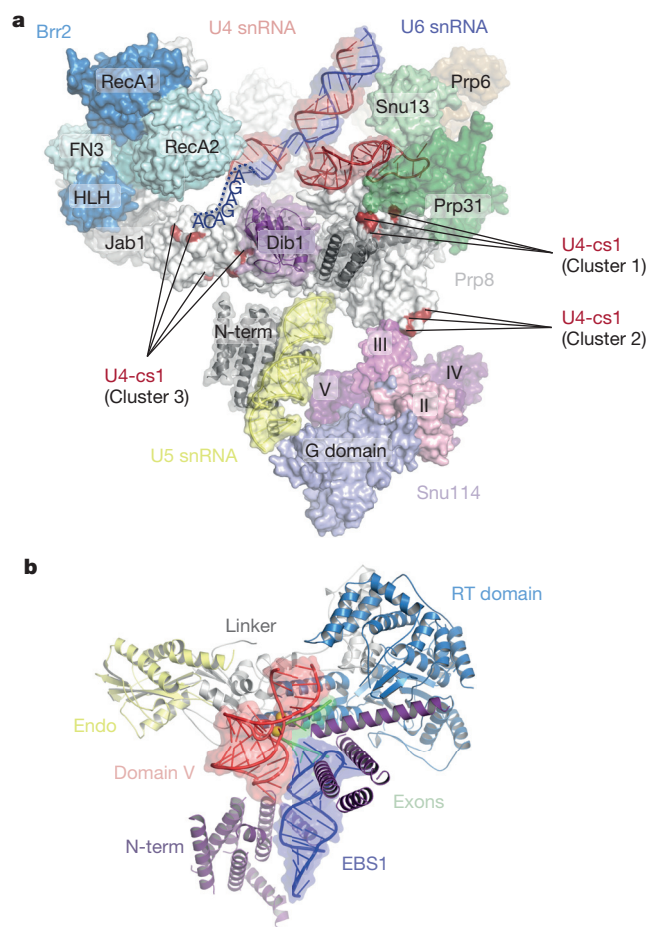
### Central role of Prp8 in tri-snRNP assembly

Our single-particle cryoEM reconstruction of yeast U4/U6.U5 tri-snRNP has revealed a nearly complete organization of its RNA and protein components, although some densities remain unassigned, and Snu66, Snu23, Prp38 and possibly sub-stoichiometric Spp381 are yet to be located (Extended Data Table 1 and Extended Data Figs 4f and 8d, e). Prp8 positioned at the centre of the assembly functions as a hub of protein-protein and protein-RNA interactions, holding the whole assembly together (Figs 1 and 2b). In yeast, a stable Prp8-Snu114-Aar2-U5 core domain complex is imported into the nucleus<sup>40</sup>, where Brr2 replaces Aar2. The Jab1/MPN and RNaseH-like domains, held tightly onto the Prp8 large domain by Aar2 (Fig. 2d), are released in tri-snRNP wherein the Jab1/MPN domain forms a stable complex with Brr2 as in the crystal structure<sup>31,32</sup> (Figs 1b and 2b and Extended Data Fig. 5b). The tri-snRNP structure provides the first glimpse of the interaction between Snu114, the U5 core domain and the N-terminal domain of Prp8 which holds the co-axially stacked stems I and II, and variable stem-loop of U5 snRNA (Fig. 3). On the opposite side of Prp8, Snu13 and Prp31 firmly bound to 5' stem-loop of U4 snRNA<sup>33</sup>, the U4 Sm protein assembly, the Brr2-Jab1/MPN domain complex, the Prp3-Prp4 complex, the RNaseH domain of Prp8, and the U4/U6 snRNA duplex assemble together (Figs 1 and 2).

Prp8 has a surface on which are exposed the 5' splice-site-binding ACAGAGA sequence of U6 snRNA, the U6 sequences which pair with U2 snRNA, and U5 snRNA loop I which interacts with exon 1 and exon 2. This surface is partly occluded by a highly conserved protein, Dib1 (Figs 2c and 6), suggesting its potential role in regulating the incorporation of RNA components into the active site cavity during spliceosome assembly and activation. When U4 and U6 snRNAs are unwound, releasing U4 snRNA together with Snu13, Prp31, Prp3 and Prp4 (ref. 3) from the spliceosome, Brr2 stably bound to the Jab1/MPN domain of Prp8 is no longer held in place and could be repositioned during catalysis and spliceosome disassembly (Fig. 5 and Extended Data Fig. 8d).

### Brr2 mode of action during activation

The unwinding of the U4/U6 snRNA duplex is an essential step in spliceosomal activation and is catalysed by Brr2 (refs 14, 15). Like



**Figure 6 | Insights into activation mechanism and the active site of the spliceosome.** **a**, Mapping of the U4-cs1 suppressor mutations on the surface of Prp8. Three clusters of mutations are found in close proximity to the key elements of spliceosomal activation: Prp31/U4 snRNA 5' stem-loop, Snu114 and ACAGAGA box of the U6 snRNA. **b**, A model of the catalytic core of group II intron docked into the active site cavity by superposition of the EBS1 stem of group II intron (PDB 3IGI) and stem I of the U5 snRNA.

other Ski2-like helicases, Brr2 unwinds any RNA duplex with 3' overhangs<sup>41</sup>. The U4/U6 snRNA duplex has 3' overhangs on both ends (Fig. 3a and Extended Data Fig. 6a) and it has been suggested that Brr2 binds to the single-stranded region of U4 snRNA and translocates along U4 snRNA<sup>41,42</sup>. Our structure shows that Brr2 is pre-loaded onto the single-stranded region between U4 snRNA 3' stem-loop and stem I of the U4/U6 duplex, showing definitively that it translocates along U4 snRNA. The gap between the two RecA domains is widened in tri-snRNP, and the prominent separator  $\beta$ -hairpin is located adjacent to stem I of the U4/U6 snRNA duplex (Fig. 5b, c). Our purified U4/U6.U5 tri-snRNP disintegrates upon addition of ATP regardless of the presence of GTP or GDP but remains intact after addition of ADP or AMPPNP, a non-hydrolysable ATP analogue (Extended Data Fig. 8). This shows that Brr2 is in an active state in our purified U4/U6.U5 tri-snRNP, in perfect agreement with the structure. *In vitro* the RNaseH domain binds to the forked region of the U4/U6 snRNA duplex adjacent to stem I and inhibits Brr2 binding to the substrate RNA<sup>41</sup>. In tri-snRNP the RNaseH domain fails to prevent substrate loading onto Brr2 or unwinding. The helix-loop-helix domain of the Brr2 N-terminal helicase cassette interacts with the 3' stem-loop of U4 snRNA, and this interaction may be important for positioning U4 snRNA in the Brr2 active site<sup>41,42</sup>. Because of the limited resolution we cannot model the single-stranded region of U4 snRNA *de novo*. However, when we superpose the N-terminal

helicase cassette of Brr2 onto the Hel308 structure with partially unwound DNA duplex<sup>43</sup>, ten nucleotides of the substrate DNA coincide with the extra density in the Brr2 active site and six additional nucleotides can be accommodated in the density extending further to the 5' end of the U4 snRNA 3' stem-loop (Fig. 5b).

### Role of Snu114 in spliceosome activation

Snu114 shows substantial sequence similarity to EF2 (Fig. 4 and Extended Data Fig. 7), suggesting that it might induce conformational change in the spliceosome upon GTP binding or hydrolysis and regulate spliceosomal activation<sup>11–13</sup>. EF-G, the bacterial counterpart of EF2, enters the ribosome in the GTP-bound form. Its GTPase is activated when switch regions I and II are remodelled upon interacting with the sarcin–ricin loop of 23S rRNA<sup>44</sup> (Fig. 4e) and GTP hydrolysis leads to translocation<sup>45</sup>. The activation process of the spliceosome has not been dissected in detail and it is not known at what stage GTP is hydrolysed or how Snu114 GTPase is activated. Snu114 and EF2 share highly similar switch I and II sequences, including the critical His residue, which in EF-G places a water molecule adjacent to the  $\gamma$ -phosphate. Unassigned density connecting the junction between stems I and II of U5 snRNA and the switch I and II loops coincides with the position of the sarcin–ricin loop (Fig. 4d). This is likely to be the N-terminal domain of Prp8, which may have a role in the activation of GTP hydrolysis.

Before the unwinding of the U4/U6 duplex, the 5' splice site sequence pairs with the ACAGAGA sequence in U6 snRNA<sup>2</sup>. The U4-cs1 cold-sensitive mutation, which extends U4/U6 stem I at the restrictive temperature and sequesters the ACAGAGA box from the 5' splice site, stalls the spliceosome before unwinding<sup>46</sup> (Extended Data Fig. 6a). A suppressor of U4-cs1 has a duplication of the ACAGAGA sequence in U6 snRNA<sup>47</sup>. This shows that pairing of the ACAGAGA sequence with the 5' splice site is a checkpoint to ensure proper assembly of complex B before the unwinding of the U4/U6 snRNA duplex. Notably, suppressors of U4-cs1 in Prp8 form three clusters on the surface of the large domain of Prp8 (Fig. 6a and Extended Data Table 2)<sup>10</sup>. In tri-snRNP, one of these clusters is located at the interface between the RT domain and domain III of Snu114, and another is at the interface with Prp31 and the junction between the RT and N-terminal domains of Prp8, showing that this checkpoint can be bypassed when these subunit interfaces are tampered with (Fig. 6a). This suggests that the interactions between these components undergo allosteric changes, which possibly couple the guanine-nucleotide binding state of Snu114 and the pairing between 5' splice site and the ACAGAGA sequence to the activation of the U4/U6 duplex unwinding. Understanding the activation process will require extensive interplay between structural and biochemical work—the tri-snRNP structure provides an important structural framework for further investigation of this process.

The structural resemblance between the group II intron active site<sup>48</sup> and the catalytic RNA core of the spliceosome<sup>4,49</sup> endorsed the hypothesis that they evolved from a common ancestor<sup>50</sup>. On the basis of the similarity of the domain architecture between the group II intron encoded protein (IEP) and Prp8, we proposed that Prp8 evolved from IEP and recruited more domains and interacting proteins to assemble spliceosomal snRNAs<sup>10</sup>, which derived from fragmented group II intron<sup>50</sup>. We placed the catalytic core of group II intron RNA in the tri-snRNP structure by superimposing its exon binding stem-loop<sup>48</sup> onto stem I of U5 snRNA. The group II intron catalytic core fits neatly into the active site cavity after removal of Dib1, which is absent from activated spliceosomes (Fig. 6b), and with small rearrangement it can make contacts with the thumb/linker region of Prp8, which crosslinks with the catalytic RNA core in the spliceosome<sup>9,10</sup>. The structure of tri-snRNP clearly illustrates how the Prp8 domains and other spliceosomal proteins come together to assemble snRNAs and insert their functional segments into the active site cavity of the spliceosome<sup>10</sup>.



**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 24 February; accepted 6 May 2015.**

**Published online 24 June 2015.**

- Will, C. L. & Lührmann, R. Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.* **3**, a003707 (2011).
- Chan, S. P. & Cheng, S. C. The Prp19-associated complex is required for specifying interactions of U5 and U6 with pre-mRNA during spliceosome activation. *J. Biol. Chem.* **280**, 31190–31199 (2005).
- Fabrizio, P. *et al.* The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. *Mol. Cell* **36**, 593–608 (2009).
- Fica, S. M. *et al.* RNA catalyses nuclear pre-mRNA splicing. *Nature* **503**, 229–234 (2013).
- Newman, A. J. & Norman, C. U5 snRNA interacts with exon sequences at 5' and 3' splice sites. *Cell* **68**, 743–754 (1992).
- Sontheimer, E. J. & Steitz, J. A. The U5 and U6 small nuclear RNAs as active site components of the spliceosome. *Science* **262**, 1989–1996 (1993).
- Stevens, S. W. *et al.* Biochemical and genetic analyses of the U5, U6, and U4/U6.U5 small nuclear ribonucleoproteins from *Saccharomyces cerevisiae*. *RNA* **7**, 1543–1553 (2001).
- Gottschalk, A. *et al.* Identification by mass spectrometry and functional analysis of novel proteins of the yeast [U4/U6.U5] tri-snRNP. *EMBO J.* **18**, 4535–4548 (1999).
- Turner, I. A., Norman, C. M., Churcher, M. J. & Newman, A. J. Dissection of Prp8 protein defines multiple interactions with crucial RNA sequences in the catalytic core of the spliceosome. *RNA* **12**, 375–386 (2006).
- Galej, W. P., Oubridge, C., Newman, A. J. & Nagai, K. Crystal structure of Prp8 reveals active site cavity of the spliceosome. *Nature* **493**, 638–643 (2013).
- Fabrizio, P., Lagerbauer, B., Lauber, J., Lane, W. S. & Lührmann, R. An evolutionarily conserved U5 snRNP-specific protein is a GTP-binding factor closely related to the ribosomal translocase EF-2. *EMBO J.* **16**, 4092–4106 (1997).
- Small, E. C., Leggett, S. R., Winans, A. A. & Staley, J. P. The EF-G-like GTPase Snu114p regulates spliceosome dynamics mediated by Brr2p, a DEXD/H box ATPase. *Mol. Cell* **23**, 389–399 (2006).
- Bartels, C., Urlaub, H., Lührmann, R. & Fabrizio P. Mutagenesis suggests several roles of Snu114p in pre-mRNA splicing. *J. Biol. Chem.* **278**, 28324–28334 (2003).
- Raghunathan, P. L. & Guthrie, C. RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEH-box splicing factor Brr2. *Curr. Biol.* **8**, 847–855 (1998).
- Lagerbauer, B., Achsel, T. & Lührmann, R. The human U5–200kD DEXH-box protein unwinds U4/U6 RNA duplexes *in vitro*. *Proc. Natl Acad. Sci. USA* **95**, 4188–4192 (1998).
- Liu, S., Rauhut, R., Vornlocher, H. P. & Lührmann, R. The network of protein-protein interactions within the human U4/U6.U5 tri-snRNP. *RNA* **12**, 1418–1430 (2006).
- van Nues, R. W. & Beggs, J. D. Functional contacts with a range of splicing proteins suggest a central role for Brr2p in the dynamic control of the order of events in spliceosomes of *Saccharomyces cerevisiae*. *Genetics* **157**, 1451–1467 (2001).
- Sander, B. *et al.* Organization of core spliceosomal components U5 snRNA loop I and U4/U6 Di-snRNP within U4/U6.U5 Tri-snRNP as revealed by electron cryomicroscopy. *Mol. Cell* **24**, 267–278 (2006).
- Häcker, I. *et al.* Localization of Prp8, Brr2, Snu114 and U4/U6 proteins in the yeast tri-snRNP by electron microscopy. *Nature Struct. Mol. Biol.* **15**, 1206–1212 (2008).
- McMullan, G. *et al.* Experimental observation of the improvement in MTF from backthinning a CMOS direct electron detector. *Ultramicroscopy* **109**, 1144–1147 (2009).
- Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature Methods* **10**, 584–590 (2013).
- Scheres, S. H. A Bayesian view on cryo-EM structure determination. *J. Mol. Biol.* **415**, 406–418 (2012).
- Bai, X. C., McMullan, G. & Scheres, S. H. How cryo-EM is revolutionizing structural biology. *Trends Biochem. Sci.* **40**, 49–57 (2015).
- Scheres, S. H. Beam-induced motion correction for sub-megadalton cryo-EM particles. *Elife* **3**, e03665 (2014).
- Scheres, S. H. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nature Methods* **9**, 853–854 (2012).
- Jørgensen, R. *et al.* Two crystal structures demonstrate large conformational changes in the eukaryotic ribosomal translocase. *Nature Struct. Biol.* **10**, 379–385 (2003).
- Grainger, R. J., Barrass, J. D., Jacquier, A., Rain, J. C. & Beggs, J. D. Physical and genetic interactions of yeast Cwc21p, an ortholog of human SRm300/SRRM2, suggest a role at the catalytic center of the spliceosome. *RNA* **15**, 2161–2173 (2009).
- Dix, I., Russell, C. S., O'Keefe, R. T., Newman, A. J. & Beggs, J. D. Protein-RNA interactions in the U5 snRNP of *Saccharomyces cerevisiae*. *RNA* **4**, 1675–1686 (1998).
- Reuter, K., Nottrott, S., Fabrizio, P., Lührmann, R. & Ficner, R. Identification, characterization and crystal structure analysis of the human spliceosomal U5 snRNP-specific 15 kD protein. *J. Mol. Biol.* **294**, 515–525 (1999).
- Maeder, C., Kutach, A. K. & Guthrie, C. ATP-dependent unwinding of U4/U6 snRNAs by the Brr2 helicase requires the C terminus of Prp8. *Nature Struct. Mol. Biol.* **16**, 42–48 (2009).
- Nguyen, T. H. D. *et al.* Structural basis of Brr2-Prp8 interactions and implications for U5 snRNP biogenesis and the spliceosome active site. *Structure* **21**, 910–919 (2013).
- Mozaffari-Jovin, S. *et al.* Inhibition of RNA helicase Brr2 by the C-terminal tail of the spliceosomal protein Prp8. *Science* **341**, 80–84 (2013).
- Liu, S. *et al.* Binding of the human Prp31 Nop domain to a composite RNA-protein platform in U4 snRNP. *Science* **316**, 115–120 (2007).
- Schultz, A., Nottrott, S., Hartmuth, K. & Lührmann, R. RNA structural requirements for the association of the spliceosomal hPrp31 protein with the U4 and U4atac small nuclear ribonucleoproteins. *J. Biol. Chem.* **281**, 28278–28286 (2006).
- Ayadi, L. *et al.* Functional and structural characterization of the Prp3 binding domain of the yeast Prp4 splicing factor. *J. Mol. Biol.* **284**, 673–687 (1998).
- Korneta, I., Magnus, M. & Bujnicki, J. M. Structural bioinformatics of the human spliceosomal proteome. *Nucleic Acids Res.* **40**, 7046–7065 (2012).
- Nottrott, S., Urlaub, H. & Lührmann, R. Hierarchical, clustered protein interactions with U4/U6 snRNA: a biochemical role for U4/U6 proteins. *EMBO J.* **21**, 5527–5538 (2002).
- Galissou, F. & Legrain, P. The biochemical defects of prp4–1 and prp6–1 yeast splicing mutants reveal that the PRP6 protein is required for the accumulation of the [U4/U6.U5] tri-snRNP. *Nucleic Acids Res.* **21**, 1555–1562 (1993).
- Makarov, E. M., Makarova, O. V., Achsel, T. & Lührmann, R. The human homologue of the yeast splicing factor prp6p contains multiple TPR elements and is stably associated with the U5 snRNP via protein-protein interactions. *J. Mol. Biol.* **298**, 567–575 (2000).
- Boon, K. L. *et al.* Prp8 mutations that cause human retinitis pigmentosa lead to a U5 snRNP maturation defect in yeast. *Nature Struct. Mol. Biol.* **14**, 1077–1083 (2007).
- Mozaffari-Jovin, S. *et al.* The Prp8 RNase H-like domain inhibits Brr2-mediated U4/U6 snRNA unwinding by blocking Brr2 loading onto the U4 snRNA. *Genes Dev.* **26**, 2422–2434 (2012).
- Hahn, D., Kudla, G., Tollervey, D. & Beggs, J. D. Brr2p-mediated conformational rearrangements in the spliceosome during activation and substrate repositioning. *Genes Dev.* **26**, 2408–2421 (2012).
- Büttner, K., Nehring, S. & Hofner, K. P. Structural basis for DNA duplex separation by a superfamily-2 helicase. *Nature Struct. Mol. Biol.* **14**, 647–652 (2007).
- Tourigny, D. S., Fernández, I. S., Kelley, A. C. & Ramakrishnan, V. Elongation factor G bound to the ribosome in an intermediate state of translocation. *Science* **340**, 1235490 (2013).
- Lin, J., Gagnon, M. G., Bulkley, D. & Steitz, T. A. Conformational changes of elongation factor G on the ribosome during tRNA translocation. *Cell* **160**, 219–227 (2015).
- Kuhn, A. N. & Brow, D. A. Suppressors of a cold-sensitive mutation in yeast U4 RNA define five domains in the splicing factor Prp8 that influence spliceosome activation. *Genetics* **155**, 1667–1682 (2000).
- Li, Z. & Brow, D. A. A spontaneous duplication in U6 spliceosomal RNA uncouples the early and late functions of the ACAGA element *in vivo*. *RNA* **2**, 879–894 (1996).
- Toor, N., Keating, K. S., Taylor, S. D. & Pyle, A. M. Crystal structure of a self-spliced group II intron. *Science* **320**, 77–82 (2008).
- Fica, S. M., Mefford, M. A., Piccirilli, J. A. & Staley, J. P. Evidence for a group II intron-like catalytic triplex in the spliceosome. *Nature Struct. Mol. Biol.* **21**, 464–471 (2013).
- Sharp, P. A. Five easy pieces. *Science* **254**, 663 (1991).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank S. Chen, G. McMullan, J. Grimmer and T. Darling for smooth running of the EM and computing facilities; P. da Fonseca, N. Unwin, I. Sanchez Fernandez, A. Amunts, P. Emsley, G. Murshudov and A. Brown for advice; A. Easter and L. Passmore for reagents; M. Skehel for mass spectrometry; and J. Li, Y. Kondo and the members of the spliceosome group for help and advice throughout the project. We are grateful to R. Henderson, D. Barford, S. Fica, P.-C. Lin and L. Strittmatter for critical reading of the manuscript. We thank V. Ramakrishnan, J. Löwe and R. Henderson for their continuing support and encouragements. T.H.D.N. was supported in part by a Herchel Smith Research Studentship. X.-c.B. was supported by a European Union Marie Curie Fellowship. The project was supported by the Medical Research Council (MC\_U105184330 to K.N. and MC\_UP\_A025\_1013 to S.H.W.S.).

**Author Contributions** T.H.D.N. developed the purification procedure for yeast tri-snRNP, prepared EM grids, collected all EM images, processed data, calculated the maps and built and fitted most of the components into the map. W.P.G. built most of the unknown components and made essential contributions to sequence analysis, homology modelling and model fitting. X.-c.B. helped T.H.D.N. with image processing and map calculation. C.G.S. guided T.H.D.N. with EM sample preparation and data collection. A.J.N. produced the Brr2 TAPS-tagged strain and contributed to the project through his knowledge of yeast spliceosome. T.H.D.N. and W.P.G. prepared all illustrations. T.H.D.N. prepared the video. S.H.W.S. carried out multi-body refinement and oversaw the EM analysis. K.N. initiated and orchestrated the project. T.H.D.N., W.P.G., A.J.N. and K.N. interpreted the results and wrote the paper with crucial contribution from all other authors.

**Author Information** 3D cryo-EM density maps of yeast U4/U6.U5 tri-snRNP have been deposited with the Electron Microscopy Database under accession number EMD-2966. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.H.D.N. ([knguyen@mrc-lmb.cam.ac.uk](mailto:knguyen@mrc-lmb.cam.ac.uk)) or K.N. ([kn@mrc-lmb.cam.ac.uk](mailto:kn@mrc-lmb.cam.ac.uk)).

## METHODS

**Statistics.** No statistical methods were used to predetermine sample size.

**Brr2-TAPS tagging for yeast U4/U6.U5 tri-snRNP purification.** Primers specific for 55 nucleotides of the C terminus and 3' UTR of *BRR2* were used to PCR-amplify the TAPS-tag cassette together with the *KanMX6* gene from pFA6a-TAPS-kanMX6, a modified version of pFA6a-TAP-kanMX6 in which the calmodulin-binding peptide tag is replaced by two tandem copies of the StrepII tag<sup>51</sup>. The PCR product was used to transform yeast strain BCY123 [*MATa pep4::HIS3 prb1::LEU2 bar1::HIS6 lys2::GAL1/10-GAL4 can1 ade2 trp1 ura3 his3 leu2-3, 112*] by homologous recombination, selecting for G418-resistance. C-terminal TAPS-tagging of *Brr2* was confirmed by PCR analysis of genomic DNA and DNA sequencing.

**Sample preparation.** The Brr2-TAPS-tagged yeast cells (72 litres) were grown in YEPD medium to OD<sub>600</sub> of 3.5, harvested and resuspended in lysis buffer (100 mM HEPES KOH pH 8.0, 200 mM KCl, 2 mM Mg(OAc)<sub>2</sub> and 10% w/v glycerol). The cells were frozen and lysed by a Freezer Mill 6870 (SPEX CertiPrep). The crude lysate was centrifuged at 45,000 r.p.m. for 1 h. The resulting supernatants were incubated with IgG sepharose overnight at 4°C. The resin was washed with TAPS wash buffer (20 mM HEPES KOH pH 7.9, 150 mM KCl, 1 mM Mg(OAc)<sub>2</sub> and 0.1% NP40) and incubated with TAPS wash buffer in the presence of TEV protease at 4°C overnight. The flow-through was collected and incubated with Streptactin resin (GE Healthcare) for 3 h. The resin was washed with TAPS wash buffer and particles were eluted with Strep elution buffer (20 mM HEPES KOH pH 7.9, 150 mM KCl, 1 mM Mg(OAc)<sub>2</sub>, 0.1% NP40, 5 mM desthiobiotin). The eluate was subsequently applied to a 10–30% v/v glycerol gradient centrifuged at 210,000g at 4°C in a SWTi60 rotor. The fractions from the gradient were analysed by SDS-PAGE for protein composition. Glycerol was removed from the peak fractions containing tri-snRNP by dialysis against B150 buffer (20 mM HEPES KOH pH 7.9, 150 mM KCl, 1 mM Mg(OAc)<sub>2</sub>) before EM sample preparation (Extended Data Fig. 1a and 1b).

**Electron microscopy.** For cryo-EM analysis, 3.5 µl of the tri-snRNP sample was applied to Quantifoil R2/2 or R1.2/1.3 grids which were previously coated with a 6-nm-thick layer of homemade carbon film and glow-discharged (Extended Data Fig. 1c). The grids were blotted for 2 s at 4°C and plunged into liquid ethane using an FEI Vitrobot MKIII. The grids were loaded onto a Tecnai F30 Polara transmission electron microscope operated at 300 kV. Images were collected manually in low-dose mode at a calibrated magnification of 79,096×. The micrographs were recorded on either a Falcon II or a (ultra back-thinned) Falcon III detector at the same calibrated pixel size of 1.77 Å in movie mode at 17 frames s<sup>-1</sup>. A total dose of 40e<sup>-</sup> Å<sup>-2</sup> over 2.5 s, and a defocus range of 2–4 µm were used.

**Data processing.** Most steps of data processing were performed in RELION<sup>22</sup> unless otherwise stated. The 42 movie frames for each micrograph were corrected for whole-image drift using MOTIONCORR<sup>21</sup>, and contrast transfer function (CTF) parameters were estimated from the resulting micrographs using CTFFIND3 and CTFFIND4 (ref. 52). A subset of 5,000 particles was picked manually and extracted with a 280<sup>2</sup> pixel box, followed by reference-free 2D classification to obtain initial 2D class averages, which were then used as references for automatic particle picking. Particles resulting from the first round of automatic picking were extracted with a 280<sup>2</sup> pixel box for reference-free 2D classification to obtain better references for the next round of automatic particle picking. All particles from the second round of automatic particle picking were manually checked before extracting them for reference-free 2D classification (Extended Data Fig. 1d, e). Prior to both autopicking runs, the templates were low-pass filtered to 20 Å to prevent high-resolution noise bias. A total of 347,241 particles from 2,035 micrographs were selected from good 2D classes, and these particles were used for subsequent 3D processing (Extended Data Fig. 2).

A subset of 18,000 particles from only the best 2D classes of one of the data sets was used for *ab initio* 3D reconstruction by SIMPLE-PRIME<sup>53</sup> to obtain an initial model of the complex, which was low-pass filtered to 60 Å for 3D classification. 3D classification with four classes was run for 25 iterations, using an angular sampling of 7.5° and a regularization parameter *T* of 4. This resulted in two classes with much better reconstructed features than the others. These classes were combined into a subset of 179,079 particles. Auto-refinement of these particles resulted in a 7.6 Å reconstruction. These particles were subsequently used for particle-based beam-induced movement correction. For these calculations, we only used the first 30 frames of each micrograph with running averages of seven frames and a standard deviation of 1 pixel for translational alignments. We used the new 'particle-polishing' approach, which fits linear tracks through the optimal translations for all running averages and takes into account the movements of neighbouring particles on the micrographs, to further improve the accuracy of the particle-based movement corrections<sup>54</sup>. The B-factors for the resolution and

dose-dependent model for radiation damage were estimated using reconstructions from running averages of three frames.

Auto-refinement of the movement-corrected particles with a soft mask (with 12-pixel fall-off) around the entire map resulted in a map at 6.4 Å resolution while refinement with a similar soft mask around the more rigid part resulted in a map at 5.9 Å resolution (Extended Data Fig. 2 and 3a). The 5.9 Å map was used for interpretation. Our map was also validated by a tilt-pair test (Extended Data Fig. 3d). Local resolution analysis showed a wide range of resolution from 5.0 Å to 20 Å (Extended Data Fig. 4a), indicating flexibility within some parts of the structure. Further 3D classification with a finer angular sampling interval of 1.8° and local angular search range of 10° revealed conformational heterogeneity of the head, body and foot domains of the structure and did not improve the overall resolution of the map (Extended Data Fig. 2).

We used a modified refinement approach in RELION, which we term 'multi-body refinement', to improve the density for the flexible arm domain (Extended Data Fig. 3b, c). In this approach, we used masks to divide the reference map into four 'bodies', approximately corresponding to the body, head, foot and arm domains. In each iteration of the auto-refine procedure, we independently aligned every experimental particle image against projections from the four distinct bodies. To minimize errors in these alignments, before the alignment against a given body, we subtracted projections from the other three bodies from the experimental particle. Because we assume that the four bodies may adopt different relative orientations in each particle, we kept track of the most likely orientation for each of the four bodies for every particle during the course of the refinement. Thereby, subtraction of the other bodies should become ever more accurate, and this resulted in four sets of relative orientations for each particle. To express our expectation that the relative movements between the different bodies were limited, we used only local orientation searches ( $\pm 22.5^\circ$ ) in the multi-body refinement, and centred the local searches around the orientations determined for the unmasked auto-refinement mentioned above. Details of this methodology will be described elsewhere (S.H.W.S., unpublished results), and its implementation will be made available through incorporation into RELION.

All refinements used gold standard Fourier shell correlation (FSC) calculations<sup>25</sup> and reported resolutions are based on the FSC = 0.143 criterion. The FSC curves are calculated using a soft spherical mask (Extended Data Fig. 4d). Prior to visualization, all maps were corrected for the modulation transfer function of the detector and sharpened by applying a negative B-factor.

**Local resolution analyses.** Local resolution analyses were performed by Resmap<sup>54</sup> and compared with those calculated by us for each protein/RNA component of the map (Extended Data Fig. 4a–c). For the latter calculations, FSC curves are calculated using a soft spherical mask (with a 30-pixel fall-off) around each protein/RNA component of interest. Convolution effects of the masks on the FSC curves were corrected using high-resolution noise substitution<sup>55</sup>. Resolution was estimated at FSC = 0.143. These calculations were performed for each of the following components: Prp8 large domain, Prp8 RNaseH domain, Prp8 Jab1/MPN domain, Prp8 N-terminal domain, Brr2, U4 Sm with U4 snRNA 3' stem-loop, U5 Sm with Sm site, Snu114, Dib1, Prp6, Prp3, Prp4, Snu13, Prp31, LSm, U5 snRNA and U4/U6 snRNA. Extended Data Fig. 4d shows some representative curves from these calculations.

FSC curves of model versus map were calculated using the Xmipp package<sup>56</sup> and the reported resolutions were based on the FSC = 0.25 criterion. FSC curves of model versus map were calculated for not only the entire model of all components but also different parts of the maps. The map of each modelled component was extracted from the tri-snRNP map using a soft mask (with a 5-pixel soft edge) surrounding the component. A map of each model was created by the program pdb2mrc within the EMAN package<sup>57</sup>. Some proteins/domains that are close together were grouped together for these calculations, including Prp8 N-terminal domain/Dib1, Brr2/Prp8-Jab1/MPN domain and Prp3/Prp4. Extended Data Fig. 4e shows some representative curves from these calculations.

**Model fitting and building.** Locations of available X-ray or homology models were fitted initially by visual inspection of the tri-snRNP map and low-pass filtered maps (to 10 Å) were generated for each model in Chimera followed by fit optimization in Chimera<sup>58</sup>. For the LSm proteins that are in the flexible arm region, the map resulting from multi-body refinement (Extended Data Fig. 3c and 5l) was used for fitting. Further rigid-body fitting was performed in Coot<sup>59</sup>. The homology model for Snu114 was prepared by I-TASSER web server<sup>60</sup> based on the crystal structure of the yeast elongation factor 2 (ref. 26; PDB 1N0V) (Fig. 4 and Extended Data Fig. 5c). The model was manually inspected and the disordered regions were removed. The model for the ferredoxin-like domain of Prp3 is available at the Yeast Genome Center (<http://www.yeastrc.org>), which contains its structure predictions<sup>61</sup>. The model with the highest Mammoth Confidence Metric (MCM) score was selected for fitting. For Prp4, the protein sequence was input into Robetta Beta Full-chain Protein Structure (<http://robeta.bakerlab.org>), which

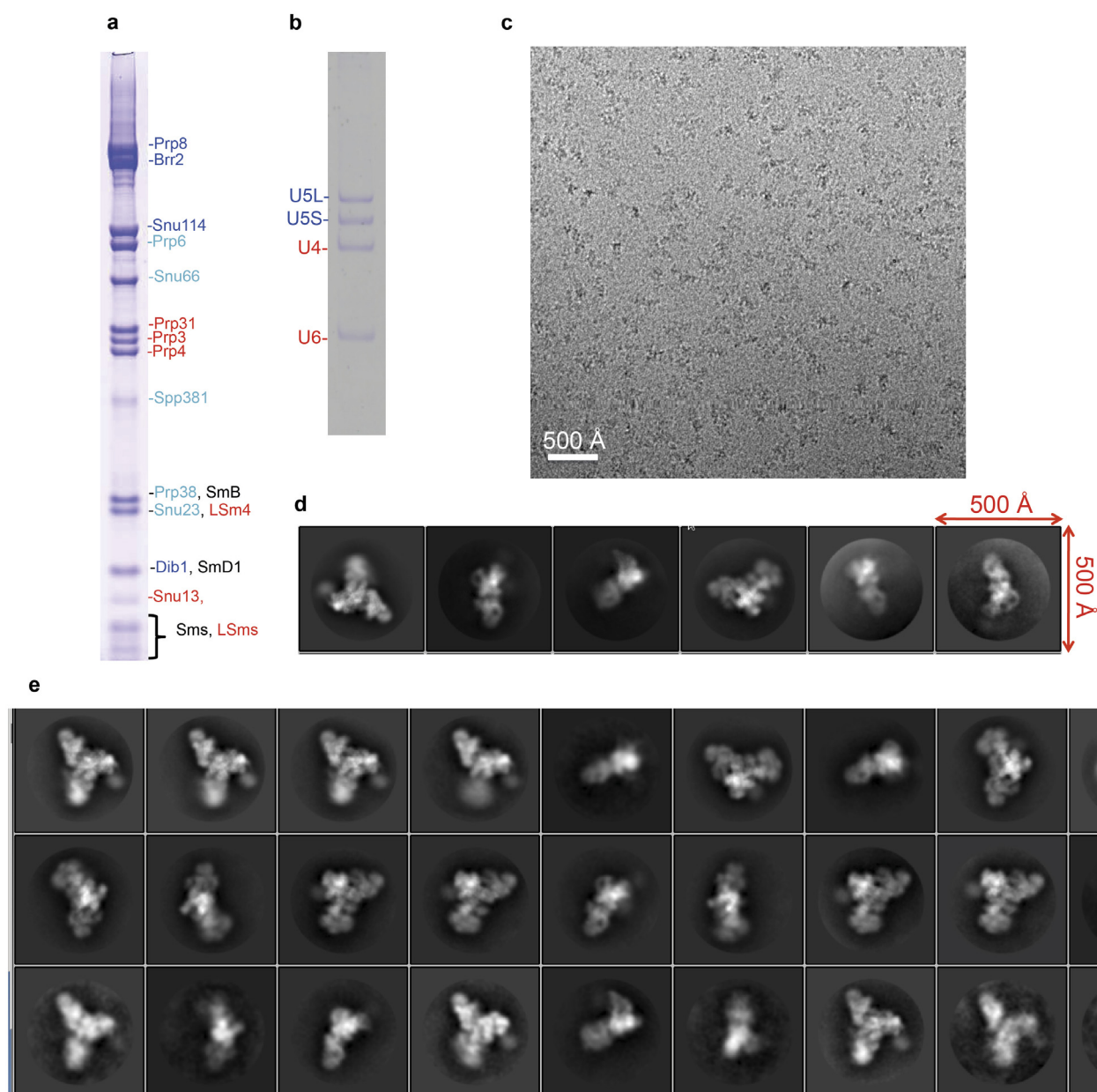


yielded a model for the C-terminal part of Prp4 based on the structure of the WDR5 protein<sup>62</sup> (PDB 3MXX). Double-stranded RNA helices and idealized polyalanine helices were built into the masked map in Coot when possible. U4 snRNA 5' stem-loop was modelled based on the structure of the human Prp31–15.5K–U4 snRNA complex<sup>33</sup> (PDB 2OZB) using ModeRNA modelling tool<sup>63</sup>. Yeast Snu13 structure<sup>64</sup> (PDB 2ALE) was fitted into the map. U4 snRNA 3' stem-loop partial model was adapted from the structure of the human U4 snRNP core domain<sup>65</sup> (PDB 4WZJ). The short and long forms of U5 snRNA<sup>66</sup> are present in our sample (Extended Data Fig. 1b) but no density for 3' stem-loop was observed, presumably because 3' stem-loop attached to the Sm site with a long single-stranded stretch is disordered or the particle population with the long U5 snRNA is classified out during classification. U5 snRNP Sm core with only the Sm site was also adapted from the human U4 snRNP core domain. The LSm proteins<sup>67</sup> (PDB 4M77) were placed in the low-resolution arm region of the map with the flat surface of the LSm complex facing the entrance side of U6 snRNA. The register of the LSm proteins cannot be accurately determined. Human Dab1 structure<sup>29</sup> (PDB 1QGV) was used for fitting. Extended Data Table 1 and Extended Data Fig. 4f summarize all the details of tri-snRNP components and modelling in our study. The active site cavity of Prp8 was described previously<sup>10</sup> and defined by crosslinks with crucial elements of U5 snRNA, U6 snRNA and pre-mRNA<sup>9</sup> and suppressors of defective splice site mutations<sup>68–74</sup>. U4-cs1 mutants have been described<sup>46,75</sup>. Extended Data Table 2 summarizes all the U4-cs1 mutants and their locations in tri-snRNP.

**Map and model visualization.** Maps were visualized in Chimera<sup>58</sup>. Map segmentation was performed in Chimera using each of the fitted models and the 'zone-masking' function (Fig. 1 and Extended Data Figs 5 and 6b, d). The LSm protein density was obtained from multi-body refinement and low-pass filtered to 20 Å. For all the remaining components, the sharpened tri-snRNP map ( $B = -214 \text{ Å}^2$ ) low-pass filtered to 5.9 Å was used. Figures were generated using either Chimera<sup>58</sup> or PyMOL (<http://www.pymol.org>) and the video was made in Chimera<sup>58</sup>.

**ATP assays.** Purified tri-snRNP from glycerol gradient (~25 nM) was incubated at 30°C for 30 min in the presence of either no nucleotide or with each of the following nucleotide combinations: ATP, ATP/GTP, ATP/GDP, ADP and AMPPNP (1 mM each). The samples (10 µl) were loaded onto a native agarose gel (0.5% in TB buffer supplemented with 1 mM MgCl<sub>2</sub>) and run at 75 V at 4°C for 2.5 h. The gel was stained with ethidium bromide for 1 h before being imaged by a Syngene UV imager (Extended Data Fig. 8a). For negative staining, the sample was also treated similarly and stained with 2% uranyl acetate before imaging on a Tecnai T12 transmission electron microscope operated at 120 kV (Extended Data Fig. 8b, c).

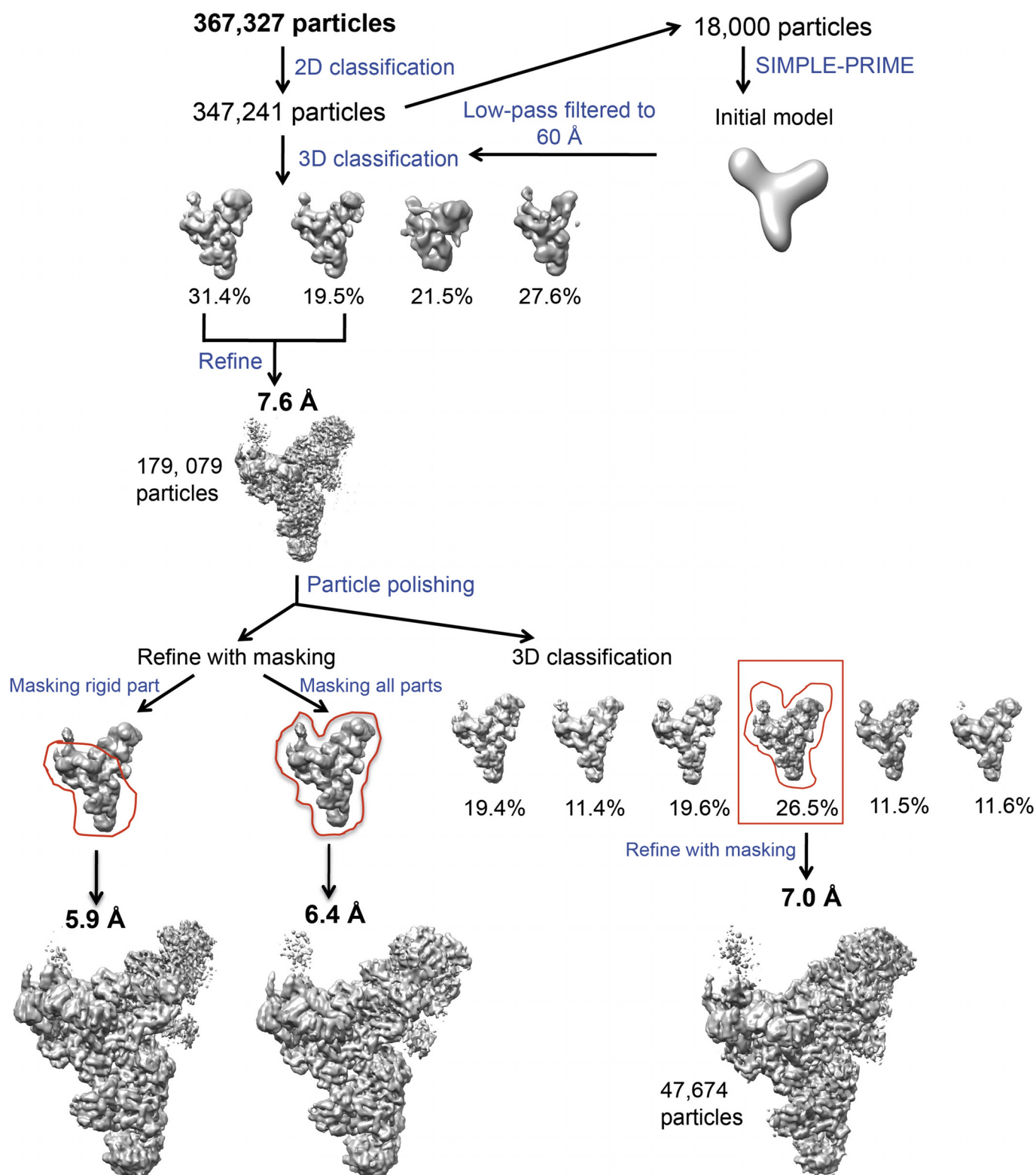
51. Schrieck, A. *et al.* RNA polymerase II termination involves C-terminal domain tyrosine dephosphorylation by CPF subunit Glc7. *Nature Struct. Mol. Biol.* **21**, 175–179 (2014).
52. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347 (2003).
53. Elmlund, H., Elmlund, D. & Bengio, S. PRIME: probabilistic initial 3D model generation for single-particle cryo-electron microscopy. *Structure* **21**, 1299–1306 (2013).
54. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nature Methods* **11**, 63–65 (2014).
55. Chen, S. *et al.* High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* **135**, 24–35 (2013).
56. Scheres, S. H. W., Nuñez-Ramírez, R., Sorzano, C. O. S., Carazo, J. M. & Marabini, R. Image processing for electron microscopy single-particle analysis using Xmipp. *Nature Protocols* **3**, 977–990 (2008).
57. Tang, G. *et al.* EMAN2: an extensive image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).
58. Goddard, T. D., Huang, C. C. & Ferrin, T. E. Visualizing density maps with UCSF Chimera. *J. Struct. Biol.* **157**, 281–287 (2007).
59. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
60. Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**, 40 (2008).
61. Malmström, L. *et al.* Superfamily assignments for the yeast proteome through integration of structure prediction with the gene ontology. *PLoS Biol.* **5**, e76 (2007).
62. Wu, X. H., Chen, R. C., Gao, Y. & Wu, Y. D. The effect of Asp-His-Ser/Thr-Trp tetrad on the thermostability of WD40-repeat proteins. *Biochemistry* **49**, 10237–10245 (2010).
63. Rother, M. *et al.* ModeRNA server: an online tool for modeling RNA 3D structures. *Bioinformatics* **27**, 2441–2442 (2011).
64. Dobbyn, H. C. *et al.* Analysis of pre-mRNA and pre-rRNA processing factor Snu13p structure and mutants. *Biochem. Biophys. Res. Commun.* **360**, 857–862 (2007).
65. Leung, A. K., Nagai, K. & Li, J. Structure of the spliceosomal U4 snRNP core domain and its implication for snRNP biogenesis. *Nature* **473**, 536–539 (2011).
66. Chanfreau, G., Elela, S. A., Ares, M. Jr & Guthrie, C. Alternative 3'-end processing of U5 snRNA by RNase III. *Genes Dev.* **11**, 2741–2751 (1997).
67. Zhou, L. *et al.* Crystal structures of the Lsm complex bound to the 3' end sequence of U6 small nuclear RNA. *Nature* **506**, 116–120 (2014).
68. Query, C. C. & Konarska, M. M. Suppression of multiple substrate mutations by spliceosomal *prp8* alleles suggests functional correlations with ribosomal ambiguity mutants. *Mol. Cell* **14**, 343–354 (2004).
69. Umen, J. G. & Guthrie, C. Mutagenesis of the yeast gene *PRP8* reveals domains governing the specificity and fidelity of 3' splice site selection. *Genetics* **143**, 723–739 (1996).
70. Liu, L., Query, C. C. & Konarska, M. M. Opposing classes of *prp8* alleles modulate the transition between the catalytic steps of pre-mRNA splicing. *Nature Struct. Mol. Biol.* **14**, 519–526 (2007).
71. Dagher, S. F. & Fu, X. D. Evidence for a role of Sky1p-mediated phosphorylation in 3' splice site recognition involving both Prp8 and Prp17/Slu4. *RNA* **7**, 1284–1297 (2001).
72. Ben-Yehuda, S. *et al.* Extensive genetic interactions between *PRP8* and *PRP17/CDC40*, two yeast genes involved in pre-mRNA splicing and cell cycle progression. *Genetics* **154**, 61–71 (2000).
73. Collins, C. A. & Guthrie, C. Allele-specific genetic interactions between Prp8 and RNA active site residues suggest a function for Prp8 at the catalytic core of the spliceosome. *Genes Dev.* **13**, 1970–1982 (1999).
74. Siatecka, M., Reyes, J. L. & Konarska, M. M. Functional interactions of Prp8 with both splice sites at the spliceosomal catalytic center. *Genes Dev.* **13**, 1983–1993 (1999).
75. Kuhn, A. N., Li, Z. & Brow, D. A. Splicing factor Prp8 governs U4/U6 RNA unwinding during activation of the spliceosome. *Mol. Cell* **3**, 65–75 (1999).



**Extended Data Figure 1 | U4/U6.U5 tri-snRNP sample used for this study.**  
**a**, Coomassie-blue-stained SDS-PAGE gel showing protein composition of the purified tri-snRNP. U5-, U4/U6- and tri-snRNP-specific proteins are labelled in blue, red and teal, respectively. Sm proteins present in both U5 and U4/U6 are in black. **b**, Toluidine-blue-stained denaturing acrylamide (9%)

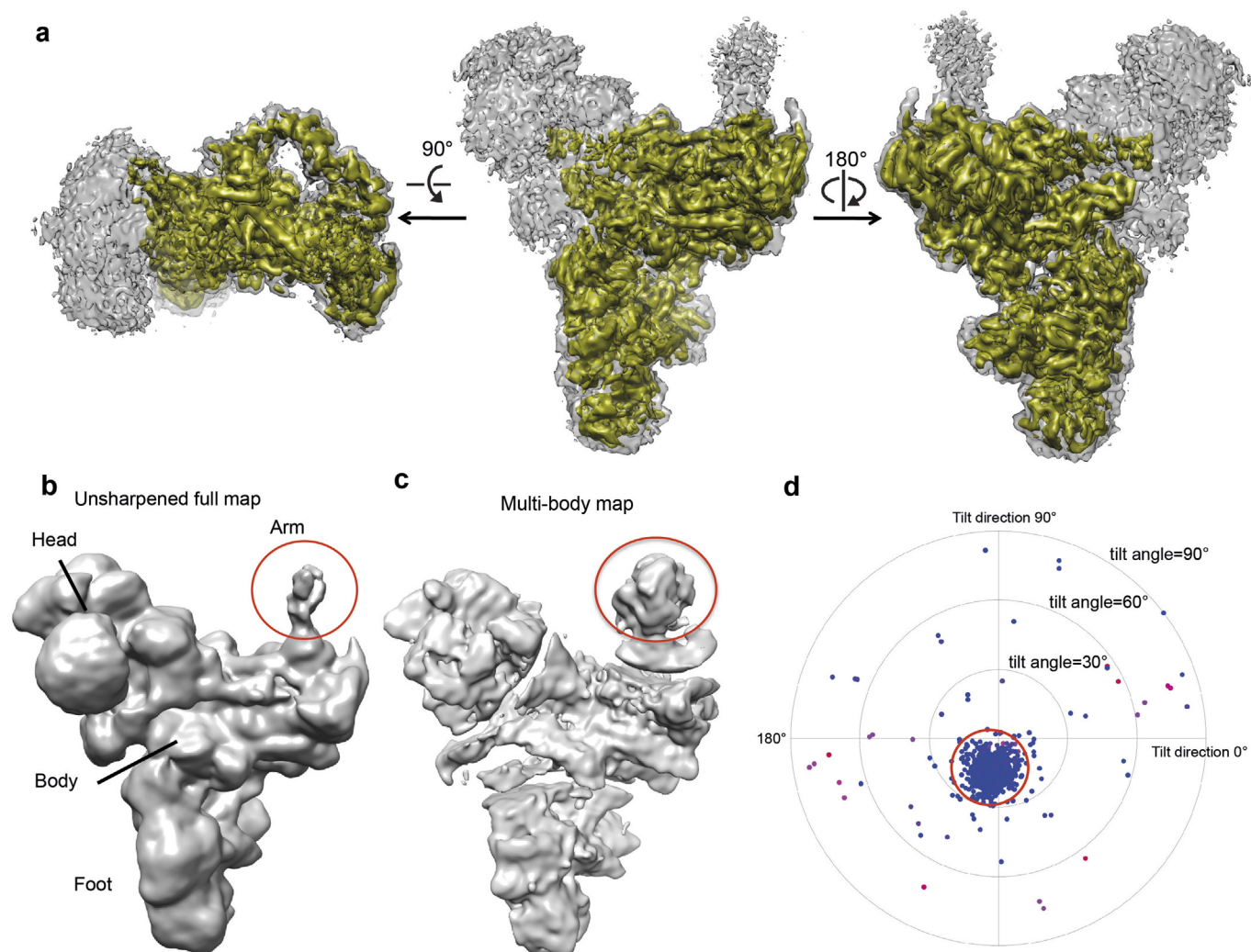
gel showing RNA compositions. **c**, Electron cryo-micrograph of tri-snRNP where the carbon-coated grid was discharged in *N*-amylamine. **d**, **e**, Reference-free two-dimensional class averages of a data set collected on a grid discharged in air and *N*-amylamine, respectively.





**Extended Data Figure 2 | Classification and refinement procedures used in this study.** A total of 367,327 particles were subjected to reference-free 2D classification. A subset of 347,241 particles from good 2D classes was selected for 3D classification using an initial model obtained from SIMPLE-PRIME<sup>53</sup>, which was low-pass filtered to 60 Å. The data were divided into four 3D classes, two of which (a total of 179,079 particles) showed better features and were combined for refinement. This resulted in a 7.6 Å reconstruction. To further improve the reconstruction, these particles were subjected to beam-induced motion correction (particle polishing)<sup>24</sup>. Refinement of these polished particles

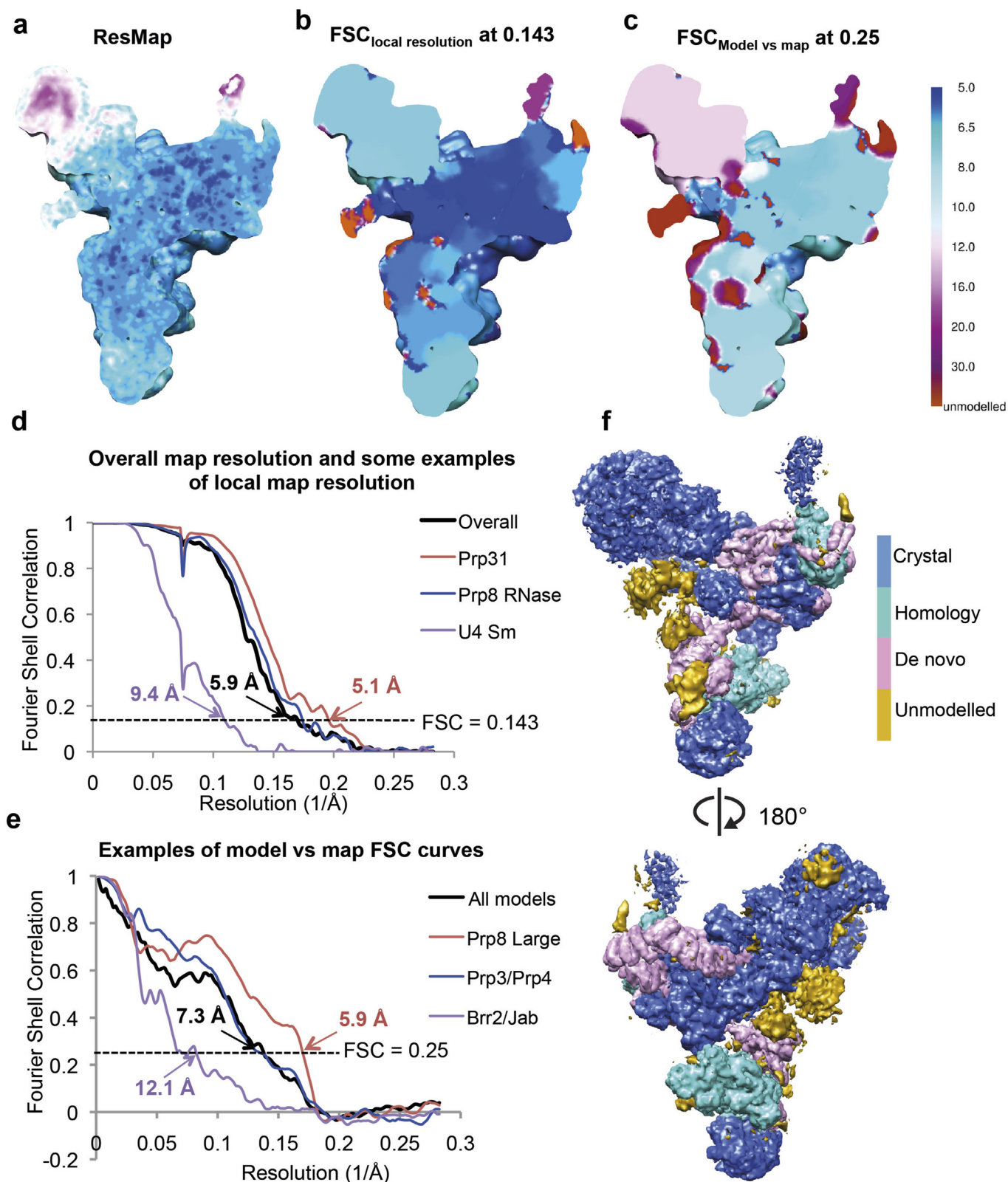
with a soft mask around the rigid part of the map (as indicated by the red envelope) yielded a 5.9 Å reconstruction while refinement with a mask around the whole map yielded a 6.4 Å reconstruction. The polished particles were also subject to further 3D classification with a finer angular sampling of 1.8°. The most populated class (47,674 particles), which also has the best rotational accuracy, was refined with a soft mask around the whole density. This resulted in a 7.0 Å reconstruction. In this study, the 5.9 Å reconstruction was used for subsequent biological interpretation. All steps were performed in RELION<sup>22</sup> unless otherwise stated.



**Extended Data Figure 3 | CryoEM maps and tilt-pair validation.** **a**, CryoEM density of the whole tri-snRNP at 5.9 Å resolution by 'gold standard' Fourier shell correlation (FSC) of 0.143 criterion at two different contour levels. The high contour map (gold) shows well-resolved densities for protein and RNA helices and flat densities for  $\beta$ -sheets. The low contour map (silver) shows densities for the more flexible head and arm. The map was sharpened by a B-factor of  $-214 \text{ \AA}^2$  and low-pass filtered to 5.9 Å as determined by RELION. **b**, The unsharpened full map of tri-snRNP. **c**, The map resulting from multi-body refinement, in which tri-snRNP is divided into four parts: the head, body,

arm and foot. This resulted in better density for the arm domain (indicated by red circles), which is at 20 Å resolution. **d**, Tilt-pair validation plot for tri-snRNP. This was obtained from 1,196 particles from 32 micrograph pairs, imaged at 0° and 10° tilt angles. The position of each dot represents the direction and the amount of tilting for a particle pair in polar coordinates. Blue dots correspond to in-plane tilt transformations; red and purple dots correspond to out-of-plane tilt transformations. Blue dots cluster in the same region of the plot at a tilt angle of approximately 10° as indicated by the red circle.

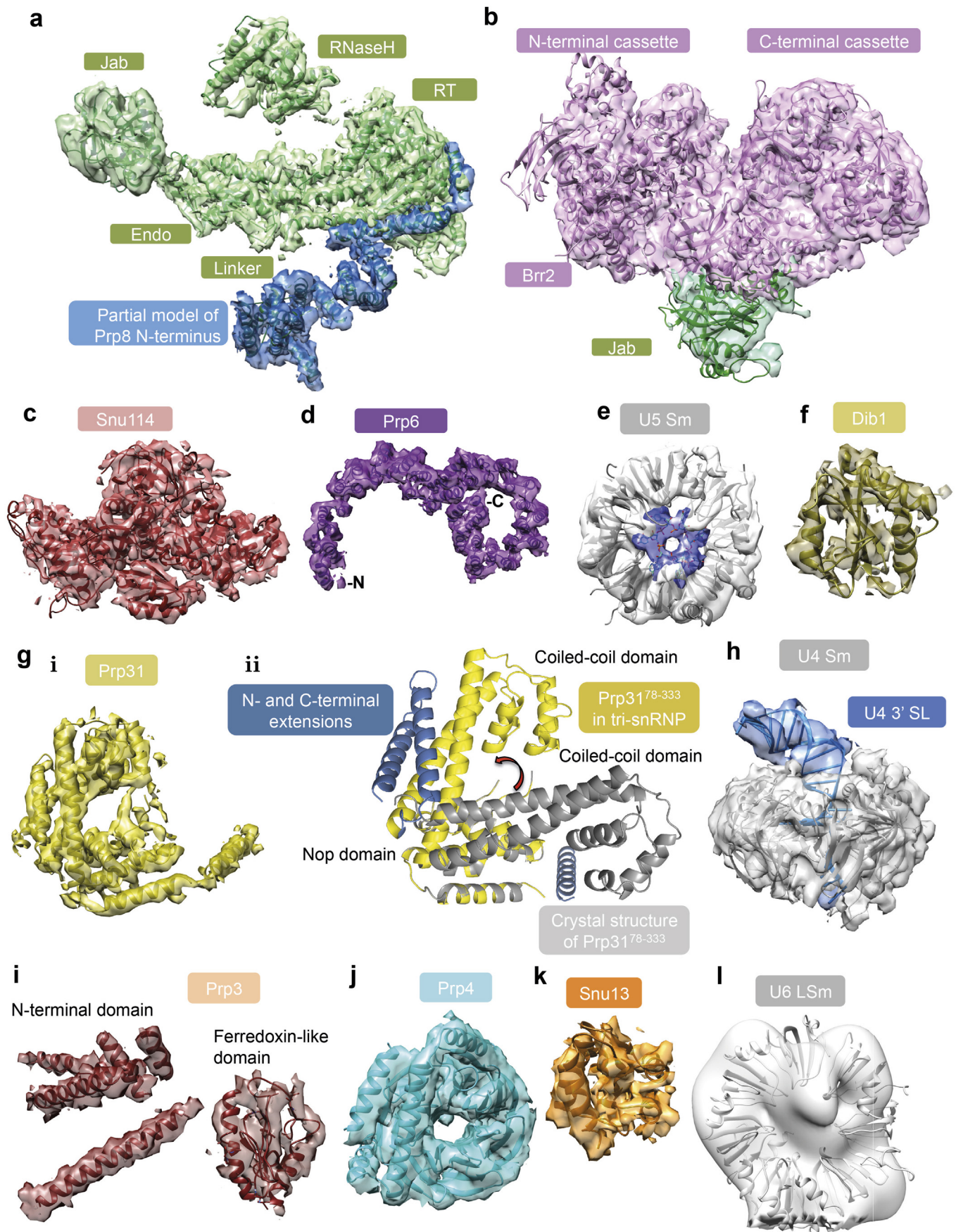




**Extended Data Figure 4 | Resolution estimation of tri-snRNP map.** **a**, Local resolution of the tri-snRNP map estimated by ResMap using the colour scheme shown in panel **c**. **b**, Local resolution of the tri-snRNP map calculated by 'gold-standard' FSC. For each component of the map that we modelled protein/RNA components, a soft mask (with a 30-pixel soft edge) surrounding the region of interest was prepared and used for FSC calculations. Convolution effects of the masks on the FSC curves were corrected using high-resolution noise substitution<sup>55</sup>. Resolution was estimated at FSC = 0.143. Local resolution for the unmodelled region of the map (in red) was not estimated. **c**, Local resolution of model versus map. The map of each modelled component was extracted from the map using a soft mask (with a 5-pixel soft edge) surrounding the component. The model was converted into density by EMAN<sup>57</sup>. FSC of

model versus map was calculated using Xmipp<sup>56</sup>. The map is coloured according to resolution estimates based on a FSC threshold of 0.25. The lower resolution estimates from the FSC of model versus map compared to the estimates from ResMap and the gold-standard FSCs are explained by the nature of our models. Because of the limited resolution of our map, we did not perform full atomic refinement, but placed known crystal structures and homology models as rigid bodies in the map. **d**, Gold-standard FSC curves for the whole tri-snRNP map and some of its components calculated as described in **b**. **e**, FSC curves of model versus map for the whole model and some of the components. **f**, The full tri-snRNP map in which portions of the structure produced from crystal structures, homology modelling and *de novo* building or unmodelled are coloured as indicated.

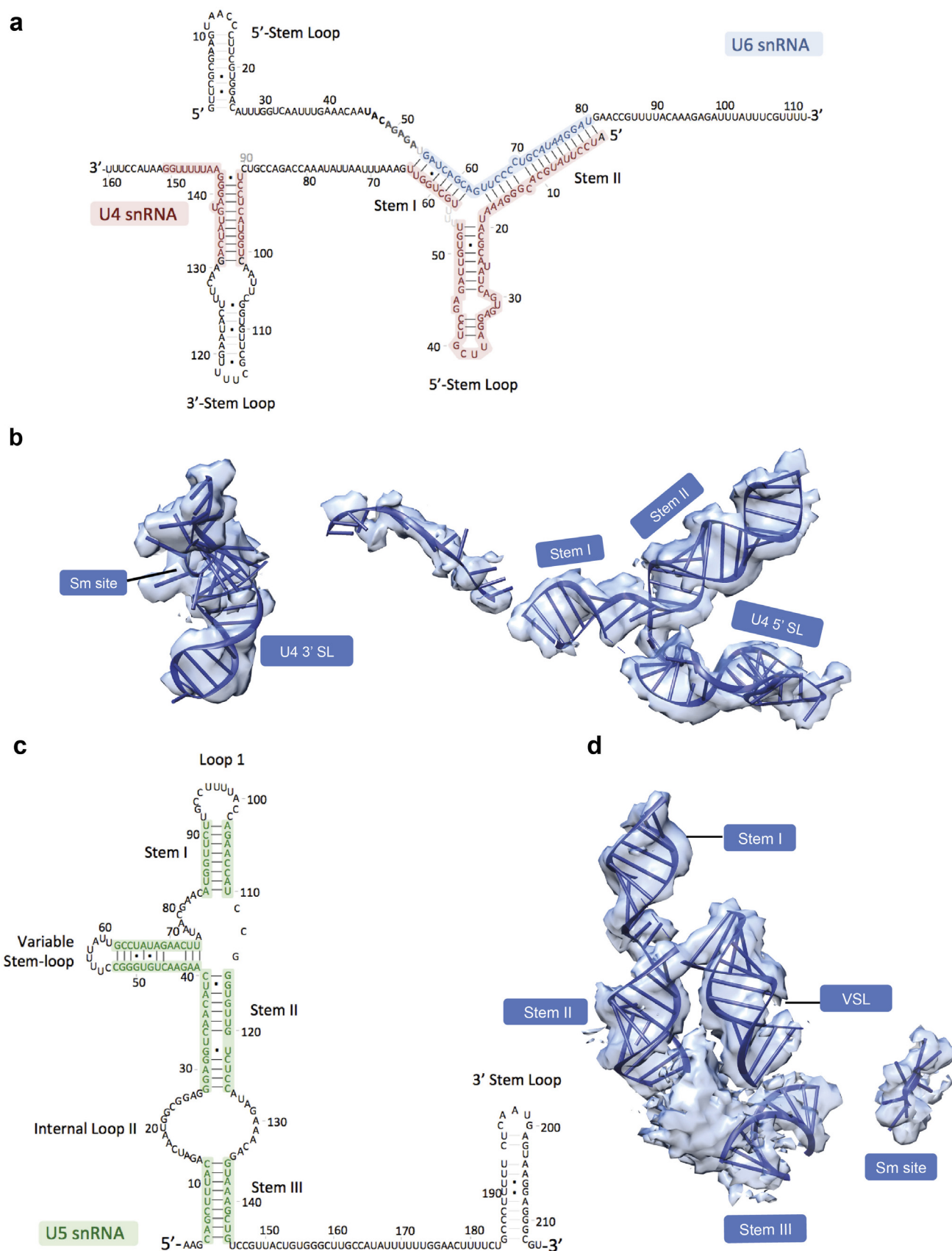




**Extended Data Figure 5 | Fitting of protein components into tri-snRNP map.** **a**, Prp8(885–2,413) crystal structure<sup>10</sup> (PDB 4I43, green) and additional helices built *de novo* assigned to the N terminus of Prp8 (blue). **b**, Brr2–Jab1/MPN complex<sup>31</sup> (PDB 4BGD). **c**, Snu114 homology model based on EF2 (ref. 26). **d**, The Prp6 TPR motifs built into the tri-snRNP map. **e**, U5 Sm proteins (grey) with Sm site (blue) based on the human U4 Sm structure (PDB 4WZJ). **f**, Dab1 (ref. 29) (PDB 1QGV). **g**, (i) Prp31. (ii), Comparison between the crystal structure of human Prp31(78–333) (ref. 33) (PDB 2OZB, grey)

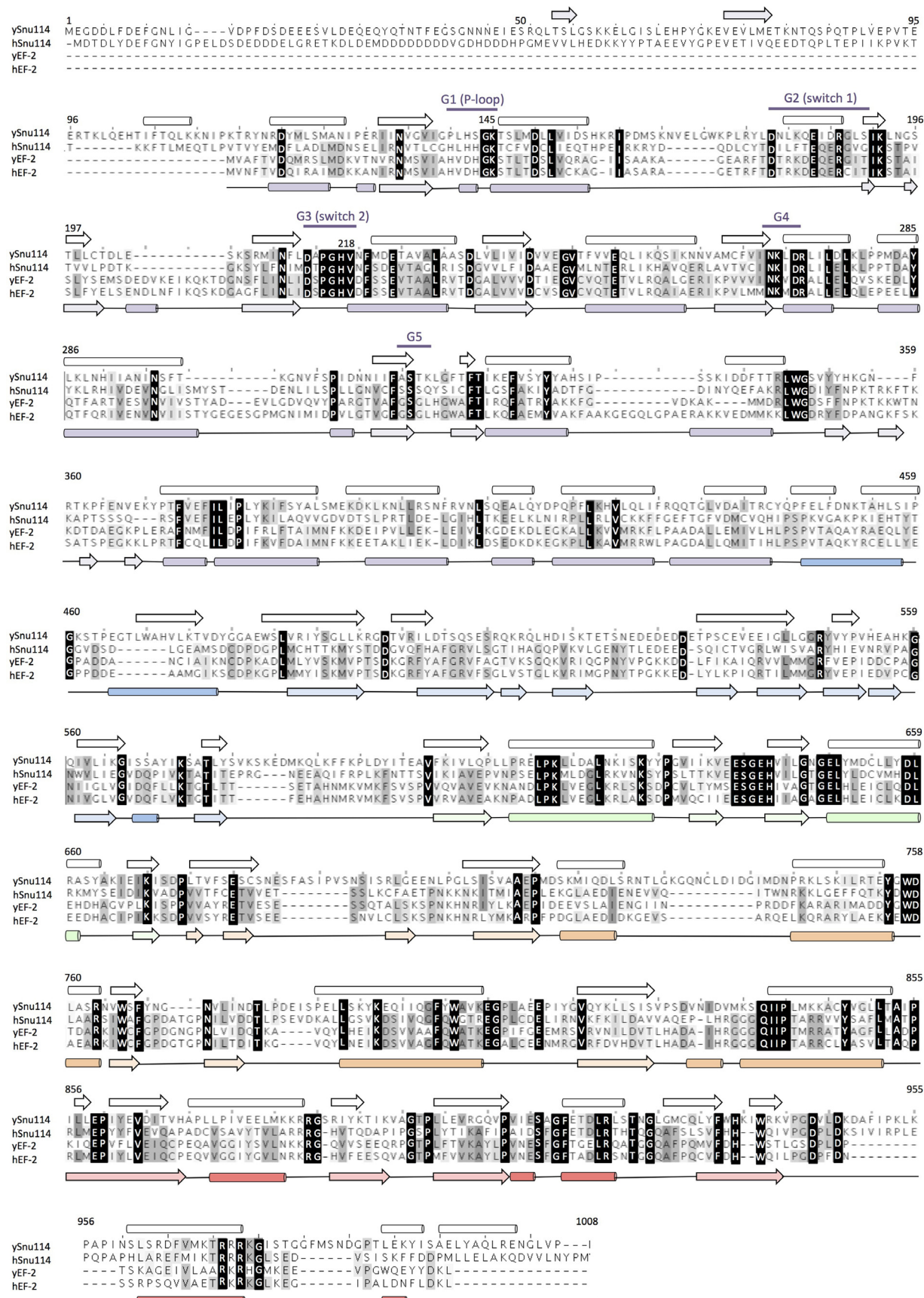
and that in tri-snRNP (yellow and blue). The coiled-coil domain (yellow) rotates by 60° in tri-snRNP with respect to the Nop domain (grey). Additional helices (blue) that extend from the N and C termini were built. **h**, U4 Sm proteins with part of U4 snRNA (blue) based on the human U4 Sm structure. **i**, Prp3 model. The ferredoxin-like domain was obtained from homology modelling while the extra helices were built *de novo*. **j**, Prp4 WD40 homology model with the extra helices built *de novo*. **k**, Snu13 (ref. 64) (PDB 2ALE). **l**, U6 LSm proteins<sup>67</sup> (PDB 4M77).





**Extended Data Figure 6 | Fitting of the RNA components in tri-snRNP map.** **a, c,** The sequences and predicted secondary structures of U4/U6 snRNA and the long version of U5 snRNA, respectively. **b, d,** The maps of the fitted

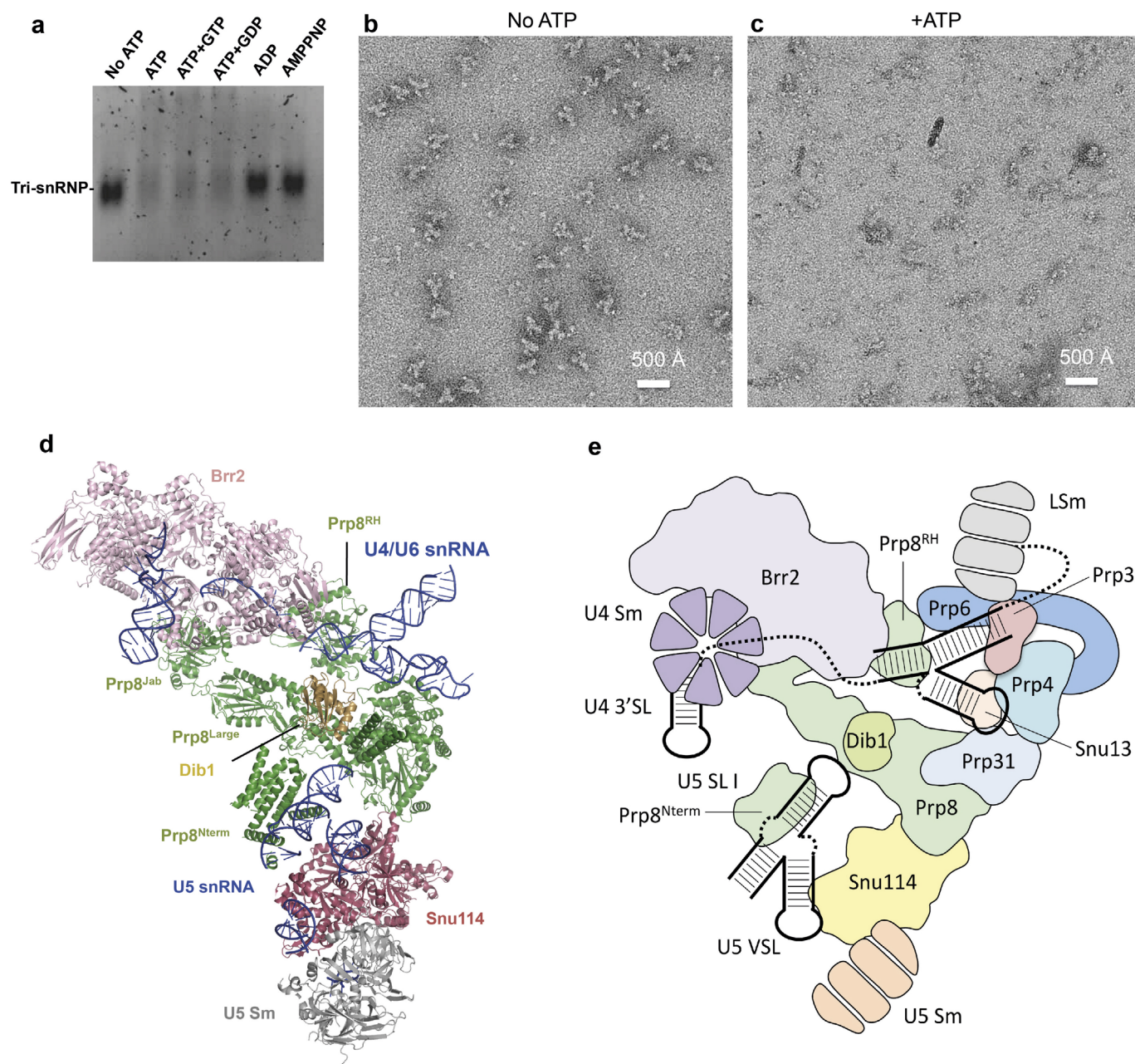
parts of U4/U6 snRNA and U5 snRNA, respectively. Unmodelled density assigned to U5 snRNA is also shown in **d**.



**Extended Data Figure 7 | Sequence alignment of yeast and human Snu114 with yeast and human elongation factor 2 (EF-2).** The secondary structures of our homology model for yeast Snu114 and the yeast EF-2 (ref. 26) (PDB 1N0V) are shown on the top and bottom of the alignment, respectively.

Important sequence elements are also shown. The greyscale shading indicates the level of sequence conservation. A higher level of conservation is shown in a darker shade.





**Extended Data Figure 8 | The effect of ATP on Brr2-TAPS purified tri-snRNP.** **a**, Ethidium-bromide-stained native agarose gel (0.5%) showing the effects of ATP addition to Brr2-TAPS purified tri-snRNP used in this study. Upon ATP addition either without or with GTP/GDP, tri-snRNP fell apart (lanes 1–4). Under the same conditions, the addition of ADP or the non-hydrolysable ATP-analogue, AMPPNP, had no effects on the complex (lanes 5, 6). **b**, **c**, The effect of ATP addition observed by negative stain microscopy. When ATP was not present, tri-snRNP particles could be observed. When ATP was added to the sample before grid preparations, tri-snRNP particles fell apart

as observed by many small components on the micrograph rather than tri-snRNP particles. **d**, Tri-snRNP model where U4/U6 snRNP proteins are not shown. In tri-snRNP, Brr2–Prp8<sup>Jab</sup> complex is loosely associated to the remaining U5 snRNP components including Prp8<sup>large</sup>, Prp8<sup>RNaseH</sup>, Prp8<sup>Nterm</sup>, Snu114, Dib1, U5 Sm proteins and U5 snRNA. After U4/U6 snRNA unwinding by Brr2, Brr2–Prp8<sup>Jab</sup> could be repositioned within the spliceosome. **e**, A schematic showing the arrangement of tri-snRNP protein and RNA components.

Extended Data Table 1 | Components and modelling of yeast U4/U6.U5 tri-snRNP

|                    | protein    | total residues | M.W.    | Domain                       | PDB code                                |
|--------------------|------------|----------------|---------|------------------------------|---|
| U5 snRNP           | Prp8       | 2413           | 279,299 | N-terminal domain            | 1-884                                   |
|                    |            |                |         | RT-like                      | 885-1251                                |
|                    |            |                |         | Thumb/X                      | 1257-1375                               |
|                    |            |                |         | Linker                       | 1376-1649                               |
|                    |            |                |         | Endonuclease                 | 1653-1824                               |
|                    | Brr2       | 2163           | 246,125 | RNaseH-like                  | 1839-1029                               |
|                    |            |                |         | Jab1/MPN                     | 2150-2396                               |
|                    |            |                |         |                              | 48GD                                    |
|                    | Snu114     | 1008           | 114,025 | N-terminal domain            | 1-441                                   |
|                    |            |                |         | G domain                     | 442-478                                 |
|                    |            |                |         | N-terminal helicase cassette | 478-1309                                |
|                    | Dib1       | 143            | 16,774  | C-terminal helicase cassette | 1330-2163                               |
|                    |            |                |         |                              | 48GD                                    |
|                    |            |                |         |                              | not modelled                            |
|                    |            |                |         |                              | not modelled                            |
|                    |            |                |         |                              | homology model (1N0V)                   |
| U4/U6 snRNP        | SmB        | 196            | 22,403  | Sm fold                      |   |
|                    | SmD3       | 110            | 11,229  | Sm fold                      |   |
|                    | SmD1       | 146            | 16,288  | Sm fold                      |   |
|                    | SmD2       | 110            | 12,856  | Sm fold                      |   |
|                    | SmE        | 94             | 10,373  | Sm fold                      | 4WZJ(human model)                       |
|                    | SmF        | 96             | 9,659   | Sm fold                      |   |
|                    | SmG        | 77             | 8,479   | Sm fold                      |   |
|                    | U5 snRNA-L | 214            | 68,847  | Loop 1                       | 92-102                                  |
|                    |            |                |         | Stem 1                       | 84-91;103-110                           |
|                    |            |                |         | IL 1                         | 75-83;111-113                           |
|                    |            |                |         | VSL                          | 41-74                                   |
|                    |            |                |         | Stem 2                       | 28-40;114-125                           |
|                    |            |                |         | IL2                          | 13-27;126-135                           |
|                    |            |                |         | Stem 3                       | 4-12;136-144                            |
|                    |            |                |         | 3' SL                        | 185-212                                 |
|                    | Snu13      | 126            | 13,570  |                              | 2ALE                                    |
|                    | Prp31      | 494            | 56,305  | N-terminal domain            | $\alpha$ -helices modelled              |
|                    |            |                |         | coiled-coil domain           | 2OZB (human model)                      |
|                    |            |                |         | Nop domain                   |   |
|                    | prp3       | 469            | 55,877  | C-terminal domain            | $\alpha$ -helices modelled              |
|                    |            |                |         | N-terminal domain            | $\alpha$ -helices modelled              |
|                    | prp4       | 465            | 52,425  | Ferredoxin-like domain       | model obtained from yeast genome center |
|                    |            |                |         |                              |   |
|                    | prp4       | 465            | 52,425  | N-terminal domain            | $\alpha$ -helices modelled              |
|                    |            |                |         | $\beta$ -propeller domain    | 166-465                                 |
|                    | SmB        | 196            | 22,403  |                              | 3MXX from Robetta prediction            |
|                    |            |                |         | Sm fold                      |   |
|                    |            |                |         | Sm fold                      |   |
|                    |            |                |         | Sm fold                      |   |
|                    |            |                |         | Sm fold                      |   |
|                    |            |                |         | Sm fold                      | 4WZJ (human model)                      |
|                    |            |                |         | Sm fold                      |   |
|                    |            |                |         | Sm fold                      |   |
|                    | LSm2       | 95             | 11,164  |                              |   |
|                    |            |                |         | Sm fold                      |   |
|                    |            |                |         | Sm fold                      |   |
|                    |            |                |         | Sm fold                      |   |
|                    |            |                |         | Sm fold                      |   |
|                    |            |                |         | Sm fold                      | 4M77                                    |
|                    |            |                |         | Sm fold                      |   |
|                    |            |                |         | Sm fold                      |   |
| tri-snRNP specific | U4 snRNA   | 160            | 51,390  | Stem I                       | 57-64                                   |
|                    |            |                |         | Stem II                      | 1-17                                    |
|                    |            |                |         | 5' SL                        | 20-53                                   |
|                    |            |                |         | central domain               | 65-80                                   |
|                    |            |                |         | 3' SL                        | 91-142                                  |
|                    |            |                |         |                              | A-form double helix                     |
|                    |            |                |         |                              | A-form double helix                     |
|                    |            |                |         |                              | homology model (2OZB)                   |
|                    | U6 snRNA   | 112            | 36,088  |                              | partial homology model (2P6R)           |
|                    |            |                |         |                              | partial model 4WZJ (human model)        |
|                    |            |                |         | 5' SL                        | 1-25                                    |
|                    |            |                |         | Stem I                       | 55-62                                   |
|                    |            |                |         | Stem II                      | 64-80                                   |
|                    |            |                |         |                              | not modelled                            |
|                    |            |                |         |                              | A-form double helix                     |
|                    |            |                |         |                              | A-form double helix                     |
| tri-snRNP specific | Prp6       | 899            | 104,234 | TPR domain                   | 1-191                                   |
|                    | Snu66      | 587            | 66,426  |                              | not modelled                            |
|                    | Prp38      | 242            | 27,957  |                              | not modelled                            |
|                    | Snu23      | 194            | 22,682  | C2H2 zinc finger-like        | not modelled                            |
|                    | Spp381     | 291            | 33,764  |                              | not modelled                            |

See Methods for details.



Extended Data Table 2 | U4-cs1 suppressors

| region    | mutations        | domains           | locations  | contact                          |
|-----------|------------------|-------------------|--|----------------------------------|
| region a  | R236G            | N-terminal domain | unknown  |                                  |
|           | L261P            |                   |  |                                  |
|           | L280P            |                   |  |                                  |
| region b  | K611R            | N-terminal domain | unknown  |                                  |
|           | E624G            |                   |  |                                  |
|           | N643S            |                   |  |                                  |
|           | V644A            |                   |  |                                  |
|           | D651G or N       |                   |  |                                  |
|           | H659P            |                   |  |                                  |
| region c  | K684E            | N-terminal domain | unknown  |                                  |
|           | E788G or V       |                   |  |                                  |
|           | N796S            |                   |  |                                  |
|           | W856R            |                   |  |                                  |
|           | E860K            |                   |  |                                  |
| cluster 1 | Q861R            | RT domain         | loop in 4 stranded $\beta$ -sheet<br>loop in 4 stranded $\beta$ -sheet<br>near in 4 stranded $\beta$ -sheet<br>near in 4 stranded $\beta$ -sheet<br>near in 4 stranded $\beta$ -sheet<br>near in 4 stranded $\beta$ -sheet | Interface with Prp31             |
|           | D1094A or N or V |                   |  |                                  |
|           | M1095T           |                   |  |                                  |
|           | V1098D           |                   |  |                                  |
|           | N1099K           |                   |  |                                  |
|           | I1104M           |                   |  |                                  |
| cluster 2 | R1105L           |                   | within loop following $\alpha$ 12<br>within loop following $\alpha$ 12<br>within loop following $\alpha$ 12  | interface with Snu114 domain III |
|           | P1191L or S or T |                   |  |                                  |
|           | D1192Y           |                   |  |                                  |
| cluster 3 | N1194D           | endonuclease      | top surface<br>top surface<br>top surface<br>top surface<br>top surface<br>side surface<br>side surface  |                                  |
|           | L1624M           |                   |  |                                  |
|           | L1634F           |                   |  |                                  |
|           | L1641F           |                   |  |                                  |
|           | T1685I           |                   |  |                                  |
|           | P1688L or R      |                   |  |                                  |
|           | A1754V           |                   |  |                                  |
| region f  | N1809D           | RNaseH            | on inner surface<br>$\beta$ -finger<br>$\beta$ -finger<br>$\beta$ -finger<br>$\beta$ -finger   |                                  |
|           | F1851L           |                   |  |                                  |
|           | V1860D or N      |                   |  |                                  |
|           | T1861P           |                   |  |                                  |
|           | V1862A or D or Y |                   |  |                                  |
|           | I1875T           |                   |  |                                  |

All suppressor mutants are described in Kuhn and Brow<sup>37</sup> and Kuhn *et al.*<sup>75</sup>.

# The core spliceosome as target and effector of non-canonical ATM signalling

Maria Tresini<sup>1</sup>, Daniël O. Warmerdam<sup>2</sup>, Petros Kolovos<sup>3</sup>, Loes Snijder<sup>1</sup>, Mischa G. Vrouwe<sup>4</sup>, Jeroen A. A. Demmers<sup>5</sup>, Wilfred F. J. van IJcken<sup>6</sup>, Frank G. Grosveld<sup>3</sup>, René H. Medema<sup>2</sup>, Jan H. J. Hoeijmakers<sup>1</sup>, Leon H. F. Mullenders<sup>4</sup>, Wim Vermeulen<sup>1</sup> & Jurgen A. Marteijn<sup>1</sup>

**In response to DNA damage, tissue homeostasis is ensured by protein networks promoting DNA repair, cell cycle arrest or apoptosis. DNA damage response signalling pathways coordinate these processes, partly by propagating gene-expression-modulating signals. DNA damage influences not only the abundance of messenger RNAs, but also their coding information through alternative splicing. Here we show that transcription-blocking DNA lesions promote chromatin displacement of late-stage spliceosomes and initiate a positive feedback loop centred on the signalling kinase ATM. We propose that initial spliceosome displacement and subsequent R-loop formation is triggered by pausing of RNA polymerase at DNA lesions. In turn, R-loops activate ATM, which signals to impede spliceosome organization further and augment ultraviolet-irradiation-triggered alternative splicing at the genome-wide level. Our findings define R-loop-dependent ATM activation by transcription-blocking lesions as an important event in the DNA damage response of non-replicating cells, and highlight a key role for spliceosome displacement in this process.**

The DNA damage response (DDR), an intricate protein network that promotes DNA repair, translesion synthesis, cell cycle arrest or apoptosis, has evolved to counteract the detrimental effects of DNA lesions<sup>1–3</sup>. At the core of the DDR, the ataxia telangiectasia mutated (ATM) and ataxia telangiectasia and Rad3-related (ATR) signalling pathways coordinate these processes in response to distinct types of DNA damage: ATR to single-stranded DNA damage, and ATM to double-strand DNA breaks (DSBs) and chromatin modifications<sup>1,4,5</sup>. These signalling networks utilize post-translational modifications and protein–protein interactions to elicit the initial stages of the cellular response. Later DDR stages involve changes in gene expression. Emerging evidence supports that DNA damage influences not only the expression levels of its target genes, by altering transcription rates and mRNA half-life, but also exon selection and ultimately their coding potential<sup>6</sup>.

Production of mature, protein-coding transcripts depends on the selective intron removal catalysed by the spliceosome, a dynamic ribonucleoprotein complex consisting of five small nuclear ribonucleoprotein (snRNP) complexes (U1, U2, U4, U5 and U6), and a large number of accessory proteins<sup>7,8</sup>. Exon/intron definition by U1 and U2 snRNPs stimulates the recruitment of the pre-assembled U4/U6.U5 tri-snRNP and numerous non-snRNP proteins. Following U1/U4 displacement and extensive conformational rearrangements, the two-step splicing reaction is catalysed by the mature, catalytically active spliceosome composed of U2, U5 and U6 snRNPs<sup>8</sup>.

The vast majority of mammalian genes are alternatively spliced to produce multiple mRNA variants from a single gene<sup>9</sup>, thus expanding protein diversity. Numerous mechanisms have evolved to provide the spliceosome with the plasticity required for selective exon inclusion, without compromising splicing fidelity<sup>9</sup>. These range from the presence of *cis*-acting elements on the transcript itself to post-translational modifications of spliceosomal proteins, which are subject to intracellular and environmental cues. Additionally, since most introns are

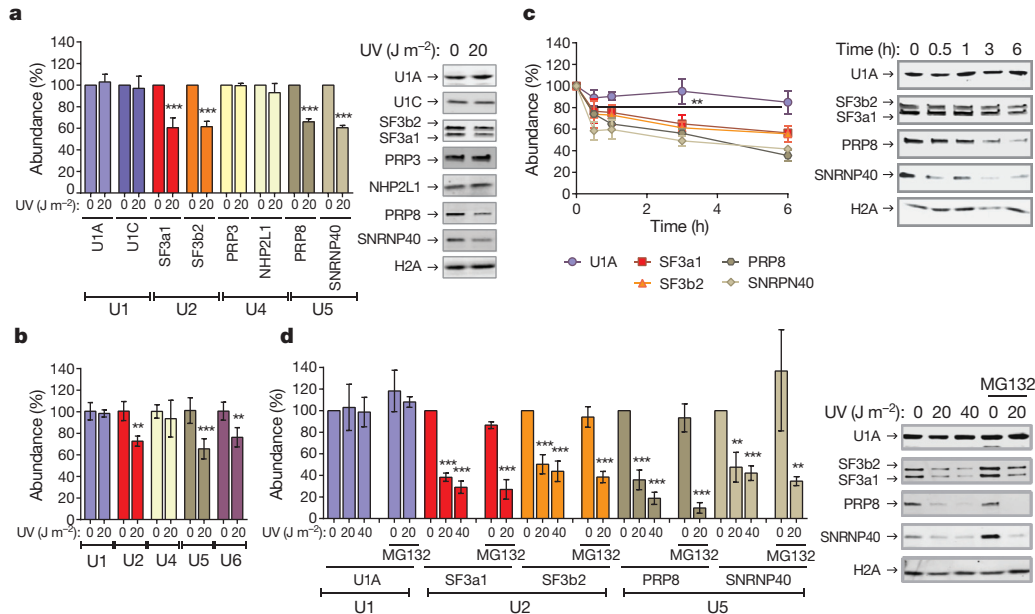
spliced co-transcriptionally within the chromatin environment, splicing decisions are subject to spatiotemporal control imposed by transcribing polymerases and interaction with chromatin remodellers and histone marks<sup>10–12</sup>. Exon selection is also influenced by DNA damage<sup>6,13</sup>. There is evidence for a broad range of damage-induced alternative splicing events, including alternative exon inclusion and exon skipping, and production of proteins with altered (often pro-apoptotic) function<sup>13–16</sup>. DNA-damage-induced alternative splicing has been attributed to changes in the processivity rate of RNA polymerase<sup>16</sup> (kinetic coupling), or changes in the interactions between the polymerase and splicing regulators<sup>14,15</sup> (recruitment coupling), under the assumption that the core spliceosome is largely unaffected. Here we present evidence that DNA damage triggers specific profound changes in spliceosome organization, primarily that of late-stage spliceosomes. Additionally, we identify reciprocal regulation between ATM-controlled DDR signalling and the core spliceosome, and show that in response to transcription-blocking DNA lesions, non-canonical ATM activation contributes to the selection of genetic information ultimately included in mature transcripts.

## DNA damage targets core spliceosomes

To gain mechanistic insight on the influence of DNA damage to chromatin-associated DDR processes, we used stable isotope labelling with amino acids in cell culture (SILAC)-based quantitative proteomic analysis<sup>17</sup> to characterize ultraviolet (UV)-irradiation-triggered changes in chromatin composition (Extended Data Fig. 1a–c). Indirect effects of replication stress were avoided by use of quiescent, human dermal fibroblasts (HDFs). UV-induced photolesions inhibit transcription by impeding RNAPII progression, and as anticipated we observed a UV-dependent chromatin-depletion of core splicing factors. Surprisingly though, this depletion was selective; chromatin abundance of all detected U2 and U5 snRNP splicing factors was

<sup>1</sup>Department of Genetics, Cancer Genomics Netherlands, Erasmus University Medical Center, Rotterdam, 3015 CN, The Netherlands. <sup>2</sup>Division of Cell Biology, Netherlands Cancer Institute, Amsterdam, 1066 CX, The Netherlands. <sup>3</sup>Department of Cell Biology, Erasmus University Medical Center, Rotterdam, 3015 CN, The Netherlands. <sup>4</sup>Department of Human Genetics, Leiden University Medical Center, Leiden, 2333 ZC, The Netherlands. <sup>5</sup>Erasmus MC Proteomics Center, Erasmus University Medical Center, Rotterdam, 3015 CN, The Netherlands. <sup>6</sup>Erasmus Center for Biomics, Erasmus University Medical Center, Rotterdam, 3015 CN, The Netherlands.





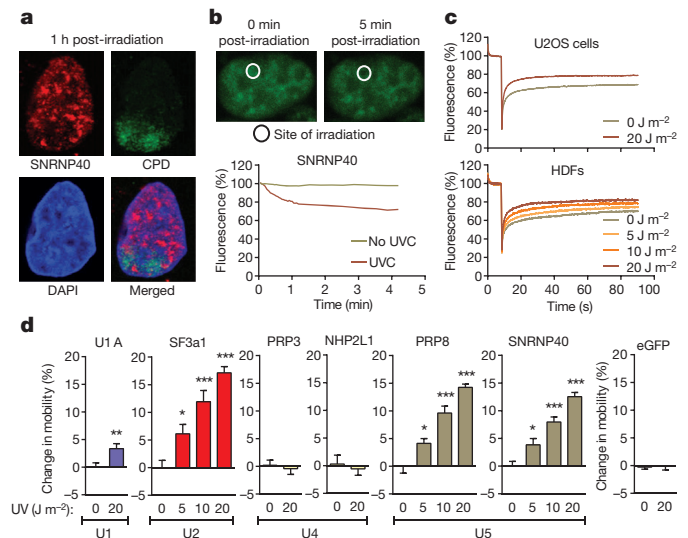
**Figure 1 | DNA-damage-triggered chromatin displacement of activated spliceosomes.** **a, b,** UV-induced changes in chromatin association of spliceosome components in quiescent HDFs. **a,** Immunoblots (right) and quantification (left) of splicing factor–chromatin association. **b,** Chromatin-associated snRNAs assayed by quantitative PCR (qPCR) and normalized to HotAir non-coding RNA ( $n = 4$ , mean  $\pm$  s.d.,  $t$ -test). **c, d,** Immunoblots (right)

and quantification (left) of splicing factor–chromatin association in U2OS cells. **c,** Time post UV irradiation. **d,** UV dose-response and lack of influence of the proteasome inhibitor MG132. Graphs in **c, d** show signal intensities normalized to H2A ( $n = 3$ , mean  $\pm$  s.d.,  $t$ -test and one-way ANOVA). \*\* $P < 0.01$ , \*\*\* $P < 0.001$ .

substantially decreased in irradiated cells while abundance of U1 and U4 snRNP splicing factors was not significantly affected (Extended Data Fig. 1d and Supplementary Table 1). Considering that spliceosomes containing exclusively U2/U5/U6 snRNPs are formed at later stages of the splicing cycle, following eviction of U1 and U4 from the assembled spliceosome<sup>8</sup>, we concluded that DNA damage preferentially targets late-maturation-stage spliceosomes, unlike chemical transcription inhibition that also affects early-stage spliceosome assembly<sup>18</sup>.

The proteomic results were validated by chromatin fractionation and immunoblotting, for U1 (U1A, U1C), U2 (SF3a1, SF3b2), U4 (PRP3, NHP2L1) and U5 (SNRNP40, PRP8) snRNP-specific proteins<sup>8</sup> (Fig. 1a). We also assayed by qPCR the chromatin association of all spliceosomal snRNAs. UV irradiation resulted in preferential chromatin depletion of U2, U5 and U6 snRNAs, while U1 and U4 were essentially unaffected (Fig. 1b). Depletion of U2 and U5 snRNP proteins was time- (Fig. 1c) and dose-dependent (Fig. 1d), but independent of proliferation status and cell type (Fig. 1a, c, d). Chromatin-depletion of U2 and U5 snRNP splicing factors was independent of proteasome activity (Fig. 1d), suggesting that depletion was not caused by splicing factor degradation but rather by relocalization. In agreement, total cellular levels of all tested splicing factors were unaffected by DNA damage (Extended Data Fig. 1e). Splicing factor relocalization was verified by immunofluorescence microscopy in cells in which DNA damage was inflicted in a small subnuclear area. A representative example in Fig. 2a depicts depletion of the U5-associated protein SNRNP40 from DNA damage sites that were identified by cyclobutane pyrimidine dimer (CPD) immunodetection<sup>19</sup>. Re-localization was monitored in real-time, using validated cell lines (Extended Data Fig. 2a–d) stably expressing GFP-tagged members of U2 (SF3a1) and U5 (SNRNP40, PRP8) snRNPs. Subnuclear damage infliction by UVC microbeam irradiation<sup>19</sup> resulted in rapid depletion from irradiated sites of GFP-tagged U2 and U5 snRNP splicing factors but not of U1 and U4 (Fig. 2b and Extended Data Fig. 3a–c). Inhibition of transcription-initiation prevented this depletion indicating that the displaced proteins were actively involved in splicing (Extended Data Fig. 3d). Irradiation of the entire cell resulted in prominent changes in splicing factor localization as evi-

denced by speckle reorganization and enlargement (Extended Data Fig. 4a, b). To further investigate the relocalization kinetics of GFP-tagged SFs, we measured their mobility by fluorescence recovery after photobleaching (FRAP). We observed substantial and UV-dose-dependent increases in the mobility of U2 and U5 snRNP factors



**Figure 2 | Mobilization and displacement of mature spliceosomes from sites of UVC-induced DNA damage.** **a,** Immunofluorescence detection of SNRNP40 and CPDs in U2OS cells exposed to UV irradiation through porous membranes. **b,** SNRNP40–GFP depletion from UVC laser microbeam irradiation sites in U2OS cells; typical image (top) and fluorescence quantification of 20 cells (bottom). **a, b,** Images were obtained at 63 $\times$  magnification. **c,** FRAP of UV-triggered SNRNP40–GFP mobilization in U2OS and quiescent HDFs ( $n = 25$ ). **d,** FRAP of free eGFP or GFP-tagged splicing factors in UV-irradiated quiescent HDFs. Change in mobility was calculated as the fluorescence of irradiated cells – fluorescence of non-irradiated cells at 1 min post-bleaching ( $n = 25$ , mean  $\pm$  s.e.m.,  $t$ -test and one-way ANOVA). \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ .

but not of U1 and U4, at 1 hour post-irradiation (Fig. 2c, d). In agreement with the chromatin fractionation assays (Fig. 1d), mobilization was independent of proteasome activity, confirming that the UV-triggered mobilization is not caused by proteasome-dependent degradation (Extended Data Fig. 5d).

The UV-dependent chromatin depletion of snRNAs and proteins participating in late-stage spliceosomes, loss of association with elongating RNAPII (Extended Data Fig. 1f), rapid displacement from DNA damage sites and mobilization of U2 and U5 snRNP factors, indicate that UV irradiation influences late-stage RNAPII-associated spliceosomes.

### DNA-damage-specific spliceosome mobilization

Next we used FRAP to address whether spliceosome mobilization is caused by specific DNA lesions or is a general response to macromolecular damage. Significant splicing factor mobilization was caused by genotoxins inflicting transcription-blocking DNA lesions (UV irradiation, Illudin S), but not oxidative damage (tert-butyl-hydroxide, rotenone, ionizing radiation), DSBs (ionizing radiation) or DNA inter-strand crosslinks (mitomycin C). This specificity argues that the observed mobilization does not result from non-specific RNA/DNA damage but only from DNA lesions that interrupt transcription (Fig. 3a and Extended Data Fig. 5a, b) and are substrates of the transcription-coupled nucleotide excision repair (TC-NER) pathway<sup>20,21</sup>. Notably, HDFs deficient in either TC-NER, or global-genome (GG)-NER (lacking CSB and XPC activities respectively), or in both (lacking XPA), show no impairments either in damage-triggered spliceosome mobilization (Fig. 3b) or in chromatin-displacement of endogenous U2 and U5 snRNP splicing factors (Extended Data Fig. 5c). Thus, the influence of transcription-blocking lesions in splicing factor localization is independent from NER complex assembly indicating that pausing of elongating RNAPII is necessary and sufficient to trigger chromatin displacement of late-stage spliceosomes.

### Spliceosome mobilization by DDR signals

Transcription inhibition by chemicals that target RNAPII mobilize splicing factors of all snRNPs, unlike UV irradiation that preferentially targets those participating in late-stage complexes (Fig. 4b

and Extended Data Fig. 5e). This preferential mobilization implies distinct mechanisms of action between UV-irradiation-dependent and chemically-induced transcription inhibition. However, to formally exclude the possibility that transcription-blocking DNA lesions mobilize spliceosomes exclusively through RNAPII arrest, we used 5,6-dichloro-1- $\beta$ -D-ribofuranosyl-benzimidazole (DRB) to inhibit transcription to the same extent as UV irradiation. Transcription arrest was evaluated by measuring reduced 5-ethynyl-uridine (SEU) incorporation into newly synthesized RNA (Fig. 4a). Both treatments increased spliceosome mobility (Fig. 4b and Extended Data Fig. 6a) and their combination had an additive effect (Extended Data Fig. 6b). Notably, UV irradiation had a more profound splicing-factor-mobilizing effect than DRB (at equal transcription-inhibiting doses), indicating that transcription inhibition alone is not sufficient to attain the extensive mobilization triggered by UV irradiation (Fig. 4b and Extended Data Fig. 6a).

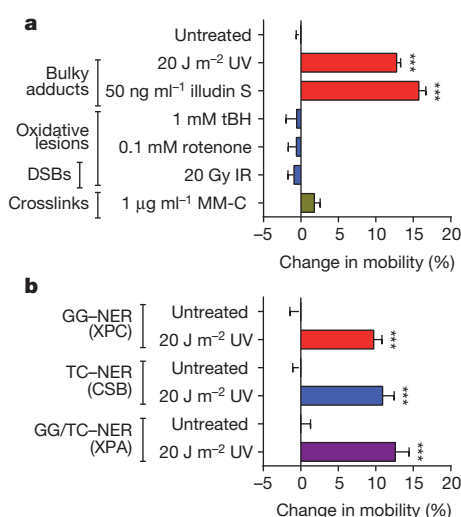
Pausing of RNAPII at DNA lesions not only halts transcription, but also activates DDR signalling pathways that modulate the cellular response via post-translational modifications<sup>1,22</sup>. Considering that many core splicing factors have been identified as DDR-kinase substrates<sup>22,23</sup>, we used the broad-range DDR-kinase inhibitor caffeine to evaluate if DDR signalling influences spliceosome organization. Caffeine partially suppressed the UV-dependent spliceosome mobilization but had no influence on the DRB-dependent mobilization, confirming that the two processes are, in part, mechanistically distinct (Fig. 4d).

To dissect which DDR signalling system augments the UV-triggered spliceosome mobilization, cells were treated with specific inhibitors of the major caffeine-sensitive DDR kinases: ATM, ATR and DNA-dependant protein kinase (DNA-PK). Neither ATR nor DNA-PK inhibition had a significant effect (Fig. 4c and Extended Data Fig. 6d). Surprisingly, ATM inhibition in non-replicating cells suppressed splicing factor mobilization to levels similar to caffeine (Fig. 4c and Extended Data Fig. 6d), while it had no influence on DRB-mediated mobilization (Extended Data Fig. 6h). The dependency of UV-triggered spliceosome mobilization on ATM signalling was confirmed by the impaired splicing factor mobilization in HDFs derived from an ataxia telangiectasia patient compared to those of a healthy donor (Fig. 4e and Extended Data Fig. 6c). Thus DNA-damage-triggered spliceosome mobilization results from the combined contribution of transcription inhibition and ATM signalling.

To evaluate the impact of ATM-dependent spliceosome mobilization on pre-mRNA processing, we assayed splicing efficiency in a select panel of DDR- and cell-cycle-related genes<sup>24</sup>. Quiescent RPE cells were UV irradiated in the absence or presence of the ATM inhibitor and intron retention was assayed by reverse-transcription PCR (RT-PCR)<sup>24</sup>. UV irradiation resulted in increased ATM-dependent intron retention (Fig. 4f and Extended Data Fig. 6f), while transcription inhibition by DRB had minor, and ATM-independent, effects. Specificity of the ATM inhibitor was confirmed by small-interfering RNA (siRNA)-mediated ATM silencing which gave identical results (Extended Data Fig. 6e).

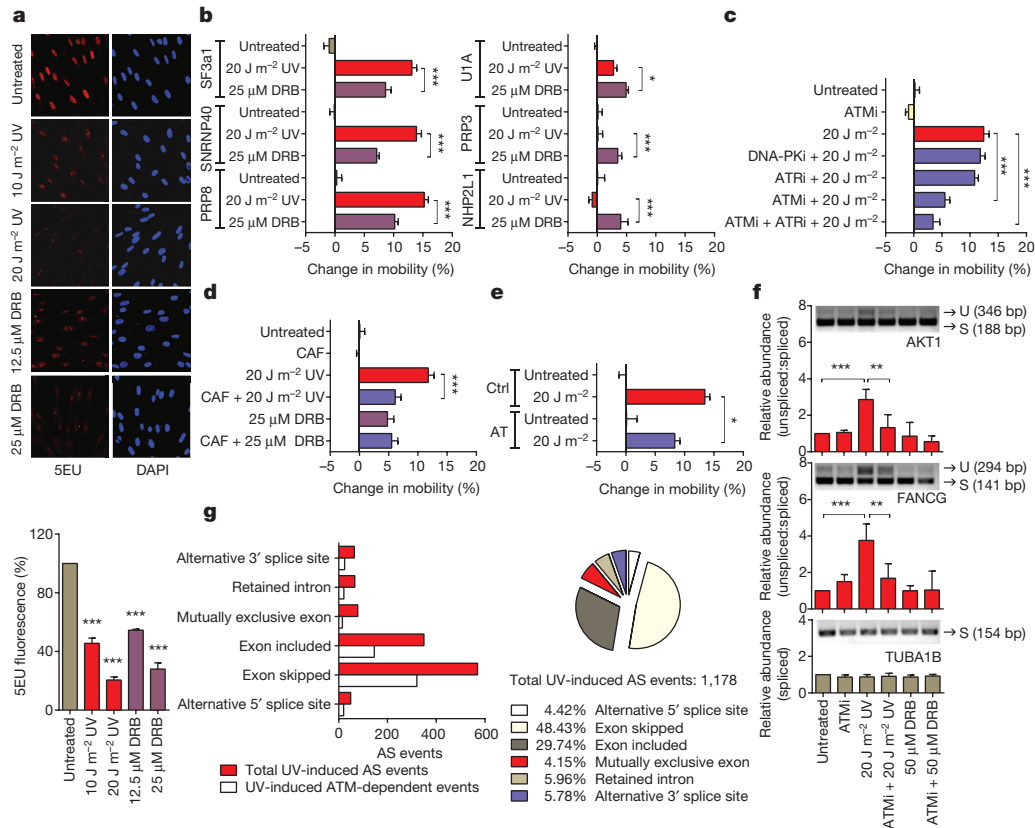
To investigate the genome-wide influence of UV irradiation on alternative splicing, as well as the ATM contribution in UV-irradiation-dependent gene expression and mRNA processing changes, we performed RNA sequencing (RNA-seq) on cells that were untreated or UV-irradiated in the presence or absence of the ATM inhibitor. We observed that a substantial number of UV-induced gene expression changes depend on ATM activity (Extended Data Fig. 6g), revealing a previously unknown contribution of ATM signalling in the UV-regulated transcriptome. Importantly, UV irradiation resulted in widespread splicing changes, a subset of which (up to 40%) was partly ATM-dependent, demonstrating the genome-wide influence of ATM not only in mRNA abundance but also in UV-induced alternative splicing (Fig. 4g and Supplementary Table 2).

Collectively, these findings demonstrate that UV irradiation influences gene expression in an ATM-dependent manner, and that ATM



**Figure 3 | Chromatin displacement of mature spliceosomes is caused by RNAPII-blocking lesions and is NER-independent.** **a**, FRAP of SNRNP40-GFP in quiescent HDFs exposed to genotoxins ( $n = 30$ , mean  $\pm$  s.e.m., one-way ANOVA). IR, ionizing radiation; tBH, tert-butylhydroperoxide; MM-C, mitomycin-C. **b**, UV-triggered mobilization of SNRNP40-GFP in HDFs deficient in GG-NER (XPC), TC-NER (CSB) or both (XPA) ( $n = 30$ , mean  $\pm$  s.e.m.,  $t$ -test). \*\*\* $P < 0.001$ .





**Figure 4 | ATM modulates spliceosome mobilization and influences splicing decisions upon DNA damage.** **a**, RNA synthesis measured by 5EU pulse labelling ( $n = 150$ , mean  $\pm$  s.e.m.,  $t$ -test). Top, representative images obtained at 20 $\times$  magnification. Bottom, graph of quantification ( $n = 150$ , mean  $\pm$  s.e.m.,  $t$ -test). **b–e**, FRAP of splicing factors in quiescent HDFs ( $n = 25$ , mean  $\pm$  s.e.m., one-way ANOVA). **b**, Response to UV or DRB treatment. **c–e**, SNRNP40 response to: **c**, UV irradiation with or without ATM, ATR or DNA-PK inhibitors (ATMi, ATRi and DNA-PKi, respectively); **d**, UV or DRB

participates in the selection of the genetic information contained in mature transcripts, thus revealing a novel non-canonical function of ATM in DDR.

### Spliceosome–ATM reciprocal regulation

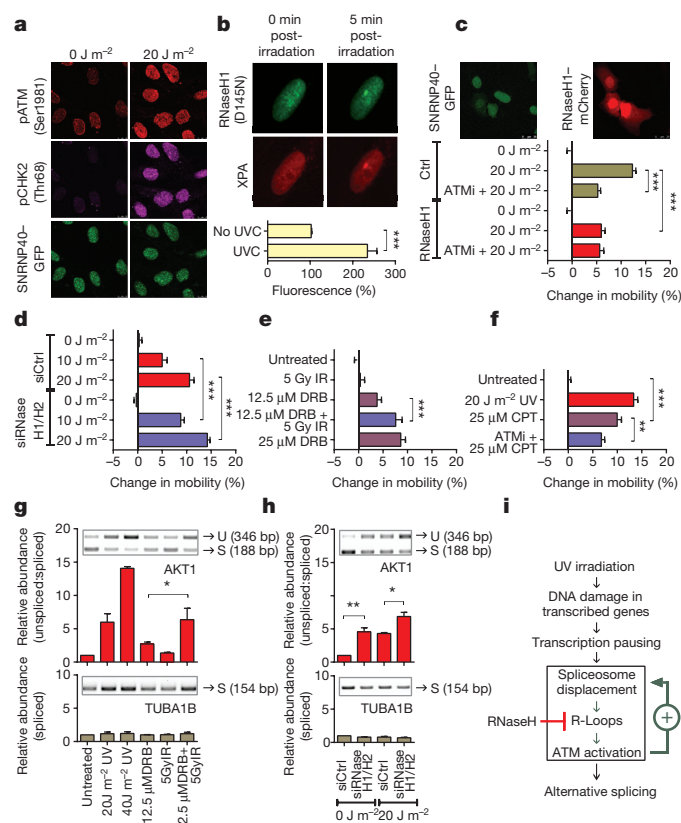
The ATM-dependency of splicing-factor-mobilization in quiescent cells indicates that UV irradiation activates ATM via a mechanism distinct from its canonical activation by replicative-stress- and ionizing-radiation-inflicted DSBs<sup>4,25</sup>. UV irradiation of quiescent HDFs activated ATM, as evidenced by its auto-phosphorylation<sup>26</sup> and phosphorylation of CHK2<sup>27</sup> (Fig. 5a and Extended Data Fig. 7a–e) to levels similar to the topoisomerase I inhibitor camptothecin (CPT)<sup>28</sup> and the deacetylase inhibitor and non-canonical ATM activator, trichostatin A (TSA)<sup>5</sup> (Extended Data Fig. 7a). Notably, in UV-irradiated cells active ATM was dispersed throughout the nucleus, which contrasts to the focal accumulation triggered by DSB-inducing agents such as CPT or ionizing radiation (Extended Data Fig. 7e). Furthermore, in cells where ATR was also inhibited<sup>29</sup>, UV-dependent  $\gamma$ H2AX and 53BP1 foci were rare (Extended Data Fig. 7f), suggesting that in non-proliferating cells UV-dependent ATM activation occurs in the absence of DSBs.

Impairments in co-transcriptional splicing promote hybridization of nascent RNA and single-stranded template DNA at the transcription bubble, resulting in three-nucleic-acid-strand structures known as R-loops<sup>25</sup>. R-loops have been reported to cause genomic instability after splicing factor depletion<sup>25,30</sup> and activate ATM in both proliferating and post-mitotic cells<sup>28,31</sup>. In agreement, siRNA-mediated silencing of

treatment with or without caffeine; **e**, UV irradiation of HDFs from an ataxia telangiectasia (AT) patient or a healthy donor (Ctrl). **f**, DRB- or UV-triggered and ATM-dependent intron inclusion assayed by RT–PCR in quiescent cells. Signal intensity expressed as unspliced:spliced ratio ( $n = 4$ , mean  $\pm$  s.d., one-way ANOVA). **g**, Genome-wide identification by RNA-seq of UV-induced alternative splicing (AS) events. Right, types of alternative splicing events; left, number of total and ATM-dependent events. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ .

U2 or U5 snRNP splicing factors, or combined RNase H1/H2A silencing, resulted in ATM activation in the absence of other treatments (Extended Data Fig. 8a, b, g). Similarly, treatment of quiescent cells with pladienolide B<sup>32</sup>, which arrests late-stage spliceosomes and mobilizes U5, and to a lesser extent U2 snRNP splicing factors (Extended Data Fig. 8c), resulted in robust ATM activation (Extended Data Fig. 8d, e) and intron-retention levels comparable to UV irradiation (Extended Data Fig. 8f). To explain our observations we formulated the following hypothesis: RNAPII arrest at DNA lesions displaces a subset of splicing factors engaged in co-transcriptional splicing. Spliceosome displacement, in combination with negative supercoiling behind RNAPII, facilitates hybridization of naked pre-mRNA (still containing intronic sequences) to the DNA template strand. The resulting R-loop activates ATM, which then amplifies the mobilization signal and stimulates further spliceosome displacement either by promoting disassembly or preventing assembly of late-stage spliceosomes. Accordingly, we predicted that: (1) R-loops are formed at sites of UV-induced DNA damage; and (2) manipulation of R-loop levels will alter spliceosome mobility.

To visualize and resolve R-loops in UV-irradiated cells we exploited the ability of RNaseH to bind and hydrolyse RNA at RNA–DNA duplexes<sup>33</sup>. For indirect, real-time visualization of R-loops, we used HDFs stably expressing GFP-tagged RNaseH1(D145N), a binding-competent but catalytically inactive RNaseH1<sup>34</sup>. RNaseH1(D145N) was rapidly recruited to UVC microbeam-irradiated sites in a transcription-dependent but ATM-independent manner (Fig. 5b and Extended Data Fig. 9d), suggesting R-loop formation at DNA-damage sites. The ability of RNaseH1(D145N) to detect R-loops was confirmed



**Figure 5 | Reciprocal regulation between spliceosome mobilization and R-loop-dependent ATM signalling.** **a**, Immunofluorescence of ATM activation in quiescent HDFs. pCHK2 and pATM, phosphorylated CHK2 and ATM, respectively. **b**, Recruitment of RNaseH1 (D145N)-GFP and mCherry-XPA at UVC microbeam irradiation sites (n = 10, mean ± s.e.m., t-test). **c-f**, FRAP showing SNRNP40-GFP mobilization in: **c**, non-transfected and mCherry-RNaseH1-expressing U2OS cells; **d**, after RNaseH1/H2A silencing; **e**, in quiescent HDFs treated with DRB and/or ionizing radiation (IR); **f**, after UV or CPT treatment. (**c-f**, n = 30, mean ± s.e.m., one-way ANOVA). **g**, **h**, Intron retention assayed by RT-PCR in quiescent cells after combined IR/DRB treatment (**g**) and silencing of RNaseH1/H2A (**h**). (**g**, **h**, n = 2, mean ± s.d., one-way ANOVA). **i**, Model of UV-triggered and R-loop/ATM-augmented spliceosome mobilization. **a**, **b**, **c**, Images were obtained at 40× (a) and 63× (b, c) magnification. \*P < 0.05, \*\*P < 0.01, \*\*\*P < 0.001.

by overexpression of active RNaseH1 or by silencing of RNaseH2, which prevented or potentiated, respectively, recruitment of RNaseH1 (D145N) at UVC microbeam-irradiation sites (Extended Data Fig. 9a). Formation of R-loops at these sites was verified using the S9.6 DNA-RNA-hybrid-specific antibody<sup>35</sup> (Extended Data Fig. 9b). Silencing of RNaseH augments R-loop abundance and resulted in detectable immunofluorescence signals at nuclear areas irradiated with doses that normally do not elicit a detectable signal, thereby confirming the S9.6 antibody specificity (Extended Data Fig. 9c).

Overexpression of active RNaseH1 attenuated the UV-induced spliceosome mobilization to levels identical to ATM inhibition (Fig. 5c and Extended Data Fig. 9e). No additional effect was observed when the two manipulations were combined (Fig. 5c), suggesting that RNaseH1 mitigates the UV-triggered spliceosome mobilization by preventing ATM activation. Conversely, silencing of RNaseH1 and H2A, which resolve the majority of RNA-DNA duplexes within the cell<sup>30,36</sup>, results in ATM activation (Extended Data Fig. 8a, b) and augments the UV-triggered R-loop formation (Extended Data Fig. 9c), spliceosome mobilization (Fig. 5d and Extended Data Fig. 9f, g) and intron retention (Fig. 5h).

ATM is required for a substantial fraction of the UV-triggered spliceosome mobilization. Regardless, ATM activation alone (for

example, as a result of ionizing radiation) does not influence spliceosome mobility (Figs 3a, 5e and Extended Data Fig. 5b), indicating that ATM controls a positive feedback mechanism that enhances, but cannot trigger, spliceosome displacement (Fig. 5i). We hypothesized that UV-dependent transcription inhibition acts as the initiating mechanism for spliceosome mobilization, which is then enhanced by a secondary ATM-dependent signal. To test this, we used treatments (DRB and ionizing radiation) that each can specifically influence one process; DRB inhibits transcription (Fig. 4a) but does not activate ATM (Extended Data Fig. 7a), while ionizing radiation activates ATM (Extended Data Figs 7a, e and 8d, e) but does not interfere with global transcription (Extended Data Fig. 5a). Combination of DRB and ionizing radiation had additive effects in both spliceosome mobilization and intron retention (Fig. 5e, g and Extended Data Fig. 10a, b), indicating that ATM amplifies (but does not initiate) a mobilization signal imposed by transcriptional arrest. In agreement, treatment of quiescent HDFs with CPT, which promotes formation of transcription-blocking lesions (Extended Data Fig. 10d) and R-loop-dependent ATM activation<sup>28</sup> (Extended Data Fig. 7e), can also efficiently mobilize spliceosomes to levels higher than expected by transcription inhibition alone (Fig. 5f and Extended Data Fig. 10c).

## Discussion

Here we present evidence that the core spliceosome is a target and an effector of the cellular response to transcription-blocking DNA damage, and we define a previously uncharacterized ATM-dependent branch of genome surveillance. Transcription-blocking DNA lesions cause selective chromatin displacement of late-stage spliceosomes by a two-step mechanism involving a stochastic (*cis*) and an ATM-signalling-mediated (*trans*) stage. Our hypothesis is that displacement of assembled co-transcriptional spliceosomes is required to remove steric inhibition that would otherwise prevent back-tracking (or removal) of RNAPII from DNA lesions, which is critical for subsequent DNA repair<sup>21</sup>. The initial spliceosome displacement probably results in naked (intron-retaining) pre-mRNA readily available for hybridization with template single-strand DNA at the transcription bubble. This culminates in R-loop formation at damaged DNA sites, which in turn activate ATM. Previously, R-loop mediated ATM activation has been linked to replication-induced DSBs because of collision of arrested transcription complexes with the replication machinery<sup>30,37</sup>. Here, we demonstrate that neither DSBs nor replication are required for R-loop-dependent ATM activation. While the exact mode of UV-triggered ATM activation remains to be determined, it does have significant biological consequences. It influences gene expression and plays a fundamental role in augmenting spliceosome displacement and alternative pre-mRNA splicing genome-wide.

ATM activation and spliceosome displacement are subject to reciprocal regulation, which has two unanticipated implications. First, in response to transcription-blocking lesions, changes in spliceosome organization activate ATM signalling irrespective of replication. Second, ATM modulates DDR, not only by controlling expression levels of its target genes, but also by influencing pre-mRNA processing. These observations provide new insights into the mechanisms and consequences of ATM activation in post-mitotic tissues, which is critical for proper cellular function, as evidenced by the severe neurodegeneration in ataxia telangiectasia patients<sup>38</sup>.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 14 April 2014; accepted 11 May 2015.

Published online 24 June 2015.

1. Sirbu, B. M. & Cortez, D. DNA damage response: three levels of DNA repair regulation. *Cold Spring Harb. Perspect. Biol.* **5**, a012724 (2013).
2. Ciccia, A. & Elledge, S. J. The DNA damage response: making it safe to play with knives. *Mol. Cell* **40**, 179–204 (2010).



3. Hoeijmakers, J. H. DNA damage, aging, and cancer. *N. Engl. J. Med.* **361**, 1475–1485 (2009).
4. Shiloh, Y. & Ziv, Y. The ATM protein kinase: regulating the cellular response to genotoxic stress, and more. *Nature Rev. Mol. Cell Biol.* **14**, 197–210 (2013).
5. Kaidi, A. & Jackson, S. P. KAT5 tyrosine phosphorylation couples chromatin sensing to ATM signalling. *Nature* **498**, 70–74 (2013).
6. Lenzken, S. C., Loffreda, A. & Barabino, S. M. RNA splicing: a new player in the DNA damage response. *Int. J. Cell Biol.* **2013**, 153634 (2013).
7. Hoskins, A. A. & Moore, M. J. The spliceosome: a flexible, reversible macromolecular machine. *Trends Biochem. Sci.* **37**, 179–188 (2012).
8. Valadkhan, S. & Jaladat, Y. The spliceosomal proteome: at the heart of the largest cellular ribonucleoprotein machine. *Proteomics* **10**, 4128–4141 (2010).
9. Kornblihtt, A. R. *et al.* Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nature Rev. Mol. Cell Biol.* **14**, 153–165 (2013).
10. Schor, I. E., Gomez Acuna, L. I. & Kornblihtt, A. R. Coupling between transcription and alternative splicing. *Cancer Treat. Res.* **158**, 1–24 (2013).
11. Zhou, H. L., Luo, G., Wise, J. A. & Lou, H. Regulation of alternative splicing by local histone modifications: potential roles for RNA-guided mechanisms. *Nucleic Acids Res.* **42**, 701–713 (2014).
12. Alexander, R. & Beggs, J. D. Cross-talk in transcription, splicing and chromatin: who makes the first call? *Biochem. Soc. Trans.* **38**, 1251–1256 (2010).
13. Dutertre, M., Sanchez, G., Barbier, J., Corcos, L. & Auboeuf, D. The emerging role of pre-messenger RNA splicing in stress responses: sending alternative messages and silent messengers. *RNA Biol.* **8**, 740–747 (2011).
14. Paronetto, M. P., Minana, B. & Valcarcel, J. The Ewing sarcoma protein regulates DNA damage-induced alternative splicing. *Mol. Cell* **43**, 353–368 (2011).
15. Dutertre, M. *et al.* Cotranscriptional exon skipping in the genotoxic stress response. *Nature Struct. Mol. Biol.* **17**, 1358–1366 (2010).
16. Muñoz, M. J. *et al.* DNA damage regulates alternative splicing through inhibition of RNA polymerase II elongation. *Cell* **137**, 708–720 (2009).
17. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnol.* **26**, 1367–1372 (2008).
18. Rino, J. *et al.* A stochastic view of spliceosome assembly and recycling in the nucleus. *PLoS Comput. Biol.* **3**, 2019–2031 (2007).
19. Dinant, C. *et al.* Activation of multiple DNA repair pathways by sub-nuclear damage induction methods. *J. Cell Sci.* **120**, 2731–2740 (2007).
20. Lagerwerf, S., Vrouwe, M. G., Overmeer, R. M., Foustieri, M. I. & Mullenders, L. H. DNA damage response and transcription. *DNA Repair (Amst.)* **10**, 743–750 (2011).
21. Vermeulen, W. & Foustieri, M. Mammalian transcription-coupled excision repair. *Cold Spring Harb. Perspect. Biol.* **5**, a012625 (2013).
22. Matsuoka, S. *et al.* ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science* **316**, 1160–1166 (2007).
23. Blasius, M. *et al.* A phospho-proteomic screen identifies substrates of the checkpoint kinase Chk1. *Genome Biol.* **12**, R78 (2011).
24. Ahn, E. Y. *et al.* SON controls cell-cycle progression by coordinated regulation of RNA splicing. *Mol. Cell* **42**, 185–198 (2011).
25. Lee, J. H. & Paull, T. T. ATM activation by DNA double-strand breaks through the Mre11-Rad50-Nbs1 complex. *Science* **308**, 551–554 (2005).
26. Bakkenist, C. J. & Kastan, M. B. DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation. *Nature* **421**, 499–506 (2003).
27. Matsuoka, S., Huang, M. & Elledge, S. J. Linkage of ATM to cell cycle regulation by the Chk2 protein kinase. *Science* **282**, 1893–1897 (1998).
28. Sordet, O. *et al.* Ataxia telangiectasia mutated activation by transcription- and topoisomerase I-induced DNA double-strand breaks. *EMBO Rep.* **10**, 887–893 (2009).
29. Hanasoge, S. & Ljungman, M. H2AX phosphorylation after UV irradiation is triggered by DNA repair intermediates and is mediated by the ATR kinase. *Carcinogenesis* **28**, 2298–2304 (2007).
30. Aguilera, A. & Garcia-Muse, T. R loops: from transcription byproducts to threats to genome stability. *Mol. Cell* **46**, 115–124 (2012).
31. Huertas, P. & Aguilera, A. Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Mol. Cell* **12**, 711–721 (2003).
32. Kotake, Y. *et al.* Splicing factor SF3b as a target of the antitumor natural product pladienolide. *Nature Chem. Biol.* **3**, 570–575 (2007).
33. Cerritelli, S. M. & Crouch, R. J. Ribonuclease H: the enzymes in eukaryotes. *FEBS J.* **276**, 1494–1505 (2009).
34. Wu, H., Lima, W. F. & Crooke, S. T. Investigating the structure of human RNase H1 by site-directed mutagenesis. *J. Biol. Chem.* **276**, 23547–23553 (2001).
35. Bhatia, V. *et al.* BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. *Nature* **511**, 362–365 (2014).
36. Reijns, M. A. M. *et al.* Enzymatic Removal of Ribonucleotides from DNA Is Essential for Mammalian Genome Integrity and Development. *Cell* **149**, 1008–1022 (2012).
37. McManus, C. J. & Graveley, B. R. RNA structure and the mechanisms of alternative splicing. *Curr. Opin. Genet. Dev.* **21**, 373–379 (2011).
38. Biton, S., Barzilai, A. & Shiloh, Y. The neurological phenotype of ataxia-telangiectasia: solving a persistent puzzle. *DNA Repair (Amst.)* **7**, 1028–1038 (2008).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We acknowledge the Optical Imaging Center of ErasmusMC for technical support; M. Reijns and A. Jackson for the S9.6 antibody; L. Marufu for technical assistance; and N. G. J. Jaspers for intellectual input. This work was funded by the Netherlands Organization for Scientific Research (NWO) ZonMW TOP Grants 912.08.031 and 912.12.132, Horizon Zenith 935.11.042, ALW 854.11.002 and 823.02.013, the Association for International Cancer Research 10-594, European Research Council Advanced Investigator Grants 233424 and 340988, and an ErasmusMC fellowship.

**Author Contributions** M.T. designed the study, performed the majority of experiments, analysed the data and authored the manuscript with contributions from W.V. and J.A.M. J.A.A.D. performed the liquid-chromatography tandem mass spectrometry analysis, L.S. assisted in fractionation/immunoblotting experiments, J.A.M. performed S9.6 antibody immunofluorescence and assisted in UVC micro-irradiation experiments, D.W. and R.H.M. performed RT-PCR splicing assays, P.K., F.G.G. and W.v.I.J. performed RNA-seq experiments, L.H.M. and M.G.V. generated RNaseH1 constructs and cell lines. L.H.M. and J.H.J.H. provided advice. All authors reviewed and commended on the manuscript.

**Author Information** RNA-seq data have been deposited in the Sequence Read Archive with accession number SRP053034. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.T. ([m.tresini@erasmusmc.nl](mailto:m.tresini@erasmusmc.nl)), W.V. ([w.vermeulen@erasmusmc.nl](mailto:w.vermeulen@erasmusmc.nl)) or J.A.M. ([j.marteijn@erasmusmc.nl](mailto:j.marteijn@erasmusmc.nl)).

## METHODS

**Materials.** Micrococcal nuclease (MNase) and all chemicals were purchased from Sigma-Aldrich unless otherwise specified. DNA-modifying enzymes were from Roche Applied Sciences. Pladienolide B was from Santa Cruz Biotechnology, the ATR inhibitor VE821 from TIB-Moltools, and the ATM inhibitor KU55933 and DNA-PK inhibitor NU7441 from R&D Systems. Antibodies used were against: PRP8 (H300), XPA/p62 (FL-273), p89/XPB (S-19) and  $\beta$ -tubulin from Santa Cruz Biotechnology; SNRPC/U1C (NBP1-96048), NHP2L1 (NBP1-32732), SF3a1 (NB100-79847), SF3b2 (NV100-79843), RNaseH1 (NBP2-20171), and RNaseH2A (NBP1-76981) from Novus Biologicals; SNRNP40 (SAB2701506) and SRSF2/SC35 (clone SC-35) from Sigma; SNRPA/U1A (3F9-1F7) from ABGENT; PRPF3 (ab187535), RNPII CTD (phospho-S2) (ab5095), RNAPII (ab5095), PCNA (PC-10), Ki67 (ab833) from abcam; CPD (TDM-2) from MBL International; GFP (11 814 460 001) from Roche; H2A (07-146) from Millipore Corp.; phospho-ATM(1981)(05-740) from Upstate Biotechnology, phospho-CHK2(Thr68) (2661) from Cell Signaling. Anti-XPC (rabbit-polyclonal ab) was in-house developed. Odyssey-compatible IRDye680- and IRDye800-conjugated secondary antibodies were from LI-COR. Secondary antibodies conjugated to Alexa Fluorochromes-488, -568, -594 and -647 were from Invitrogen. GFP-tagged proteins were immunoprecipitated with GFP-Trap beads (ChromoTek).

**Cell culture, SILAC labelling and cell treatments.** Cell lines used in this study were: ataxia telangiectasia patient (AT2)- and healthy adult donor (C5Ro)-derived human dermal fibroblasts (HDFs); SV40-transformed XPA (XP12RO), XPC (XP4A) and CSB (CS1AN) patient-derived HDFs; hTERT immortalized HDFs (C5Ro-T), VH10 human foreskin fibroblasts (VH10-T) and human retinal pigmented epithelial cells (RPE1, ATCC), human osteosarcoma cells (U2OS, ATCC); and the amphotropic retroviral packaging cell line Gryphon A (Allele Biotechnology). Cells were subcultured under standard culture conditions (37 °C, 5% CO<sub>2</sub>) in a humidified incubator. U2OS, Gryphon A, and SV40-transformed cells were grown in Dulbecco's Modified Eagle's Medium (DMEM, Lonza), supplemented with 10% v/v fetal bovine serum (FBS, Fisher Scientific) and 1% v/v penicillin-streptomycin (PS, Lonza). Primary and TERT-immortalized HDFs and RPE-1 cells were cultured in Ham's F10 (Lonza) supplemented with 15% FBS and 1% PS. When applicable, cells were synchronized in quiescence by 72 h serum-deprivation. For FRAP and immunofluorescence experiments, cells were seeded on 25-mm-diameter glass slides. For UVC laser/live-cell-imaging experiments, cells were seeded on quartz coverslips (010191T-AB; SPI Supplies). For stable isotopic labelling with amino-acids in culture (SILAC), C5Ro-T cells were cultured for >5 population doublings (PD) in lysine-, arginine- and leucine-free DMEM (AthenaES) supplemented with antibiotics, non-essential amino-acids (Lonza), 10% dialysed FBS (Invitrogen) and 105  $\mu\text{g ml}^{-1}$  leucine and either 73  $\mu\text{g ml}^{-1}$  light [<sup>12</sup>C<sub>6</sub>]lysine and 42  $\mu\text{g ml}^{-1}$  [<sup>12</sup>C<sub>6</sub>, <sup>14</sup>N<sub>4</sub>]arginine or with heavy [<sup>13</sup>C<sub>6</sub>]lysine and [<sup>13</sup>C<sub>6</sub>, <sup>15</sup>N<sub>4</sub>]arginine (Cambridge Isotope Laboratories). In each subcultivation, cell numbers were determined using a Beckman Z2 coulter counter (Beckman Coulter, Inc.), and  $0.5 \times 10^4$  cells were seeded per cm<sup>2</sup> of growth surface area. The increase in population doubling ( $\Delta\text{PD}$ ) was calculated using the formula  $\Delta\text{PD} = \log_{10}(\text{number of cells harvested}/\text{number of cells seeded})/\log_{10}(2)$ .

Cells were UVC irradiated (254 nm, TUV Lamp, Philips) at the indicated doses. For local DNA damage infliction, cells were UV-irradiated (60 J m<sup>-2</sup>) through isopore polycarbonate membranes containing 5- $\mu\text{m}$ -diameter pores (Millipore). Chemicals were added directly in the growth media at the indicated concentrations. In FRAP experiments cells were assayed 1 h after initiation of treatment with the exception of illudin S and rotenone, which were assayed at 6 h. Pre-incubation with caffeine (10 mM), DDR-kinase inhibitors (10  $\mu\text{M}$ ) and MG132 (50  $\mu\text{M}$ ), started 1 h before genotoxic treatments and lasted throughout the experiment.  $\alpha$ -Amanitin treatments were for >24 h. For exon-specific RT-PCR cells were lysed 6 h after treatment.

**Mass spectrometry and data analysis.** Nanoflow liquid chromatography-tandem mass spectrometry (LC-MS/MS) and data analysis were as described<sup>39</sup>. In brief, samples containing MNase-digested chromatin were size-fractionated by SDS-PAGE, gels were cut in 2-mm slices, and subjected to dithiothreitol-reduction, iodoacetamide alkylation and trypsin digestion. LC-MS/MS was performed on an 1100 series capillary liquid chromatography system (Agilent Technologies) coupled to an LTQ-Orbitrap XL mass spectrometer (Thermo Scientific) operating in positive mode. Raw mass spectrometry data were analysed using the MaxQuant software. A false discovery rate of 0.01 for proteins and peptides and a minimum peptide length of six amino acids were set. The Andromeda search engine was used to search MS/MS spectra against the International Protein Index (IPI) human database. Statistical analysis was performed with Perseus (1.5.0.30)<sup>17</sup>.

**Cloning.** Human full-length cDNA clones used for subcloning were; PRP8/PRPF8/DHX16 (CS116070), SF3A1 (SC321295), SNRNP40 (SC112670) and RNaseH1 (SC319446) from Origene and U1A/SNRPA (MHS6278-202826119), NHP2L1 (MHS6278-202839330) and PRP3/PRPF3 (MHS6278-202826220) from Dharmacon. To generate vectors expressing GFP- and mCherry-tagged proteins the open reading frames (minus the stop codon) of human U1A, SF3a1, PRP3, NHP2L1, PRP8 and SNRNP40 were PCR amplified using oligonucleotides containing restriction enzyme sites. PCR products were subcloned into a pLHCX retroviral expression vector (Clontech Laboratories) modified to contain eGFP lacking the initiation codon. XPA and RNaseH1 lacking the mitochondrial localization signal (amino acids 1–28) were subcloned in modified pLHCX vectors containing either eGFP or mCherry lacking their stop codons. PCR amplifications were performed on a MJ Scientific, Inc., PTC-100 Thermocycler using high-fidelity Phusion polymerase (Bioke). Amplified cDNAs were purified using the Promega Wizard kit. Following restriction digestion of inserts and vectors, shrimp alkaline phosphatase treatment of the vectors, and agarose gel electrophoresis, the gel-excised DNAs were purified using the Promega Wizard kit. DNA inserts were ligated into vectors at a 3:1 molar ratio. Plasmid DNAs were validated by restriction digestion and sequencing.

**Infections/transfections.** C5Ro-T, C5Ro, AT-2, U2OS and VH-10T cell lines stably expressing GFP-tagged proteins were generated by retroviral infection followed by hygromycin selection. For retrovirus production Gryphon A cells were transfected with the appropriate expression vector using FuGENE 6 (Roche) according to the manufacturer's instructions. Viral supernatants were harvested 48 h post-transfection, filtered through 0.45- $\mu\text{m}$  filters (Millipore Corp.) and used immediately to infect subconfluent cell cultures in the presence of 5  $\mu\text{g ml}^{-1}$  polybrene. U2OS cells were transiently transfected with RNaseH1-mCherry (pLHCX) using FuGENE 6. For gene silencing the following siRNAs were purchased from Thermo Scientific as SMARTpools: ON-TARGETplus Human RNaseH2A siRNA (L-003535-01-0005) targeting the catalytic subunit A of RNaseH2, ON-TARGETplus RNaseH1 siRNA (L-012595-01-0005), ON-TARGETplus PRP8 siRNA (L-012252-01-0005), ON-TARGETplus SF3a1 siRNA (L-016051-01-0005), ON-TARGETplus ATM siRNA (L-003201-00-00005) and a control/scrambled siRNA duplex (D-001210-05-05). For gene silencing, U2OS cells were transfected with RNAiMAX (Invitrogen), and C5RoT and RPE cells with HiPerfect (Qiagen), as recommended by the manufacturers. To inhibit endogenous RNaseH activity cells were transfected with a (1:1) mixture of siRNAs targeting RNaseH1 and RNaseH2A.

**Preparation of whole cell lysates, chromatin fractionation, and immunoprecipitations.** Whole-cell lysates were prepared by lysis of equal cell numbers in 60 mM Tris-Cl (pH 6.8), 2% SDS, 10% glycerol, 5%  $\beta$ -mercaptoethanol and 0.01% bromophenol blue. Crude chromatin was isolated after Triton-X 100 extraction and MNase digestion. All fractionation steps were performed at 4 °C. Cell pellets were suspended in a non-denaturing isosmotic buffer (10 mM PIPES (pH 7.0), 3 mM MgCl<sub>2</sub>, 100 mM NaCl, 300 mM sucrose, 0.5 mM Na<sub>2</sub>VO<sub>4</sub>, 5 mM NaF, 5 mM Na<sub>4</sub>P<sub>2</sub>O<sub>7</sub>, 10 mM  $\beta$ -glycerolphosphate, 0.1 mM PMSF, 1 mM EGTA, 1 $\times$  EDTA-free protease inhibitor cocktail (Roche), 15  $\mu\text{M}$  MG132, 10 mM N-ethylmaleimide and 20  $\mu\text{M}$  PR-619 (LifeSensors)) and extracted in the same buffer with 0.5% (v/v) Triton-X 100 for 5 min. Following centrifugation (650g, 5 min), nuclei depleted from soluble nucleoplasm were washed with MNase digestion buffer (50 mM Tris-Cl (pH 7.5), 4 mM MgCl<sub>2</sub>, 50 mM KCl, 300 mM sucrose, 0.5 mM Na<sub>2</sub>VO<sub>4</sub>, 5 mM NaF, 5 mM Na<sub>4</sub>P<sub>2</sub>O<sub>7</sub>, 10 mM  $\beta$ -glycerolphosphate, 1 mM PMSF, 1 mM EGTA and 1 $\times$  EDTA-free protease inhibitor cocktail) and subsequently incubated with 0.3 U MNase (Sigma)/1  $\times 10^6$  nuclei, and 1 mM CaCl<sub>2</sub> (37 °C, 10 min). Addition of (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> to a final concentration of 250 mM was used to facilitate extraction of stably DNA-bound proteins. EGTA and EDTA were added to 5 mM and samples were centrifuged at 16,000g for 20 min. Protein concentrations were determined using a modified Bradford method (Bio-Rad). For GFP immunoprecipitations, cells were lysed in 20 mM Tris-Cl (pH 7.5), 5 mM MgCl<sub>2</sub>, 150 mM NaCl, 0.5% Triton X-100, 1 $\times$  phosphatase inhibitor (Roche) and 1 $\times$  protease inhibitor cocktail. Chromatin was mechanically sheared by passing through a 27G syringe, 40 times. Particulate matter was removed by centrifugation (20 min at 16,000g) and supernatants containing equal amounts of proteins were used for immunoprecipitation. GFP-tagged proteins were immunoprecipitated directly or after MNase digestion which was used to cleave DNA and RNA and disrupt ternary complexes. Samples were incubated (2 h, 4 °C) with pre-equilibrated GFP-Trap coupled to agarose beads (ChromoTek), and after extensive washing (10 mM Tris-Cl (pH 7.5), 150 mM NaCl, 0.5 mM EDTA, 0.5% NP-40), immunocomplexes were dissociated from the beads by heating for 10 min at 95 °C, in 120 mM Tris-Cl (pH 6.8), 4% SDS, 20% glycerol, 10%  $\beta$ -mercaptoethanol, 0.01% bromophenol blue. For immunoprecipitation of elongating RNAPII, cells were treated and extracted as for isolation of crude chromatin with the exception that instead of MNase



digestion, chromatin was mechanically sheared. Immunoprecipitations were performed by O/N incubation with either the anti-RNAPII CTD phospho-Ser2 antibody or rabbit IgG, followed by incubation with protein A/protein G agarose beads (Upstate Biotechnology).

**Immunoblotting.** Protein samples were size-fractionated on 5–20% gradient SDS-polyacrylamide gels (BioRad) and electro-transferred onto nitrocellulose membranes using a Bio-Rad Mini-Protean electrophoresis system. Abundance of proteins of interest was assayed using antibodies at concentrations recommended by their manufacturers. Membranes were incubated with primary antibodies in Tween 20/Tris-buffered saline (20 mM Tris (pH 7.4), 150 mM NaCl, 0.1% Tween 20) containing 3% w/v non-fat dry milk or, when the pATM antibody was used, 3% BSA. Following binding of the appropriate anti-mouse or anti-rabbit Alexa Fluorochrome-conjugated secondary antibody and extensive washing, proteins of interest were visualized using the Odyssey CLx Infrared Imaging System (LI-COR Biosciences). Signal intensities were quantified using the ImageQuant TL software (GE Healthcare Life Sciences).

**RNA synthesis.** Transcription levels were determined following 2-hour incubation with ethynyluridine (EU)<sup>40</sup> added directly in the culture (serum-free) media. EU incorporation was visualized using Click-iT conjugation of AlexaFluor647 (Invitrogen) according to the manufacturer's protocol. Images were obtained using a Zeiss Axio Imager Z2 upright laser-scanning confocal microscope equipped with a 63× Plan-Apochromat 1.4 NA oil-immersion lens (Carl Zeiss Inc.). Fluorescence-signal intensities were quantified using the ImageJ software (NIH). In each experiment >150 cells per condition were analysed.

**Immunofluorescence and live-cell confocal laser-scanning microscopy.** For immunofluorescence experiments, cells were fixed with 3.7% paraformaldehyde (PFA)/PBS and permeabilized in 0.5% Triton-X 100/PBS. For detection of splicing factors and XPC, cells were pre-extracted with 0.5% Triton-X 100/PBS before fixation (0.5 min). For CPD immunodetection, nuclear DNA was denatured with 0.07 M NaOH for 5 min. For SRSF2/SC35 immunodetection cells fixed in 2% PFA/0.2% Triton-X 100/PBS were treated with 100% acetone (5 min, –20 °C). Non-specific antigens were blocked in 3% BSA/PBS. R-loop immunodetection with the S9.6 antibody was as described previously<sup>35</sup>. In brief, PFA-fixed cells were permeabilized by Triton X-100, followed by extraction with 0.5% SDS. Cells were blocked with 3% BSA, 0.1% Tween 20 in 4 × Saline Sodium Citrate (SSC) buffer. Hybridization of primary antibodies was overnight at 4 °C, and with secondary Alexa Fluorochrome-conjugated antibodies for 1 h at room temperature. Coverslips were mounted on glass slides using 4,6-diamidino-2-phenylindole (DAPI)-containing ProLong Gold antifade reagent (Molecular probes) and imaged on a Zeiss Axio Imager Z2 upright laser-scanning confocal microscope.

Live-cell-imaging experiments were performed with a Leica TCS SP5 AOBs laser scanning confocal microscope equipped with an environmental chamber (37 °C, 5% CO<sub>2</sub>). Kinetic studies of GFP-tagged proteins were performed using UVC (266 nm)-laser-irradiation for local DNA damage infliction<sup>19</sup>. In brief, a 2 mW pulsed (7.8 kHz) diode-pumped solid-state laser emitting at 266 nm (Rapp OptoElectronic) was connected to the confocal microscope with an Axiovert 200M housing adapted for UV by all-quartz optics. By focusing the UVC laser inside cell nuclei without scanning, only a limited area within the nucleus (diffraction limited spot) was irradiated. Cells were imaged and irradiated through a 100×, 1.2 NA Ultrafluor quartz objective lens. Images obtained before and after UVC laser irradiation were analysed using the LASAF software (Leica). Fluorescence intensity in the irradiated area or a non-irradiated area in the nucleus was normalized to levels in the same area before irradiation. Data were expressed as the percentage change in relative fluorescence intensity. In each experiment at least ten cells were analysed and all experiments were performed a minimum of three times.

Mobility of GFP-tagged proteins was measured by strip-FRAP as described<sup>41</sup>. In brief, a narrow (~1 µm) strip spanning the width of the nucleus was photobleached at ~20% of the initial GFP-signal intensity using a 488 nm laser at 100% power. Recovery of fluorescence in the strip was monitored at 25-ms intervals. Images obtained were analysed using the LASAF software (Leica). FRAP data were normalized to the fluorescence levels before photobleaching after subtraction of the background signal. In each experiment 8–10 cells per condition were analysed and all experiments were performed at least three times. A negative (untreated) and positive (20 J cm<sup>–2</sup> UV) control were included in all experiments.

**Chromatin-associated RNA isolation and snRNA qPCR.** Chromatin-associated RNA was isolated by a modification of the method developed by Wu and Schibler<sup>42</sup> for the isolation of ternary-complex-associated nascent RNA. Briefly, cell pellets were re-suspended in 20 mM HEPES (pH 7.5), 10 mM KCl, 250 mM sucrose, 5 mM MgCl<sub>2</sub>, 1 mM EGTA, 1 mM PMSF, 1 µl ml<sup>–1</sup> RNasin (Invitrogen), 1× phosphatase inhibitor (PhosStop, Roche) and 1× protease inhibitor cocktail (Roche), and lysed by the addition of Digitonin

to 200 µg ml<sup>–1</sup> final concentration (10 min, 4 °C). Nuclei were pelleted by centrifugation (650g, 5 min) and following re-suspension in a buffer containing 20 mM Tris-HCl (pH 7.5), 75 mM NaCl, 0.5 mM EGTA, 50% glycerol, 1 mM PMSF, 1 µl ml<sup>–1</sup> RNasin, 1× protease and 1× phosphatase inhibitors, were extracted for 10 min at 4 °C by the addition of ten volumes of a solution containing 20 mM HEPES (pH 7.6), 7.5 mM MgCl<sub>2</sub>, 0.2 mM EGTA, 300 mM NaCl, 1 M urea and 1% NP-40. Pelleted nuclei were re-suspended in Qiazol reagent (Qiagen) and RNA was isolated as recommended by the manufacturer. Following digestion with DNase I (Qiagen) RNA was cleaned up with the Qiagen RNeasy mini-kit with a second, on-column, DNase I digestion. Equal amounts of RNA from each sample were reverse-transcribed using random-hexamers and SuperScript III (Invitrogen). qPCR reactions were performed using primers complementary to human snRNAs (described and validated by Galiveti *et al.*<sup>43</sup>) or the chromatin-associated HotAir ncRNA (control for data normalization), using the IQSYBR Green Supermix (Bio-Rad) in a CFX96 Touch Real-Time PCR Detection System (Bio-Rad). Absence of contaminating genomic DNA was verified by the lack of amplified products for all sample/primer sets by inclusion of mock reverse-transcription reactions in which no enzyme was added.

**Exon-specific RT-PCR: RNA extraction, reverse transcription and PCR.** Experiments were performed as described by Ahn *et al.*<sup>24</sup>. In brief, RNA was isolated from quiescent RPE-1 cells using the RNeasy kit (QIAGEN). Equal RNA amounts from each sample were reverse transcribed using random hexamers and SuperScript III RT and cDNAs were PCR-amplified using the indicated primers and Taq DNA Polymerase (NEB). PCR products were size-fractionated by gel electrophoresis and visualized by ethidium bromide staining. Signal intensities of amplified fragments containing either the unspliced or the spliced intron were normalized to the levels of the respective fragments in untreated cells and expressed as fold change in relative abundance. All experiments were repeated a minimum of three times. Amplification of constitutive exons from TUBA1B and GAPDH were used as controls for general splicing efficiency. Primers sets used for amplifications were: FANCG (exon5–6) 5'-GGATGTCCTCCTGACAGC AT-3' and 5'-GCTGTGTACACCTGGACCAA-3'; AKT1 (exon11–12) 5'-AC AAGACGGGCACATTAAG-3' and 5'-ACCGACATCATCTCGTACA-3'; AURKA (exon9–10) 5'-AATGATTGAAGTCGGATGC-3' and 5'-TCTGGC TGGGATTATGCTTC-3'; AURKB (exon6–7) 5'-TGCAGAAGAGCTGCACAT TT-3' and 5'-TCTTCAGCTCTCCCTTGAGC-3'; TUBA1B (exon2–3) 5'-CC GGGCTGTGTTGTAGACT-3' and 5'-GATCTCCTTGCCAATGGTGT-3'; ATM (exon19–20) 5'-AAGGAGCTTCTGGAGAAGAG-3' and 5'-AACT GCTCTTGAGCATCCCTT-3'; ATR (exon33–34) 5'-AAGAGCCATCCT GGCTCTC-3' and 5'-CTACCCTGGCACTCTGCAGCC-3'.

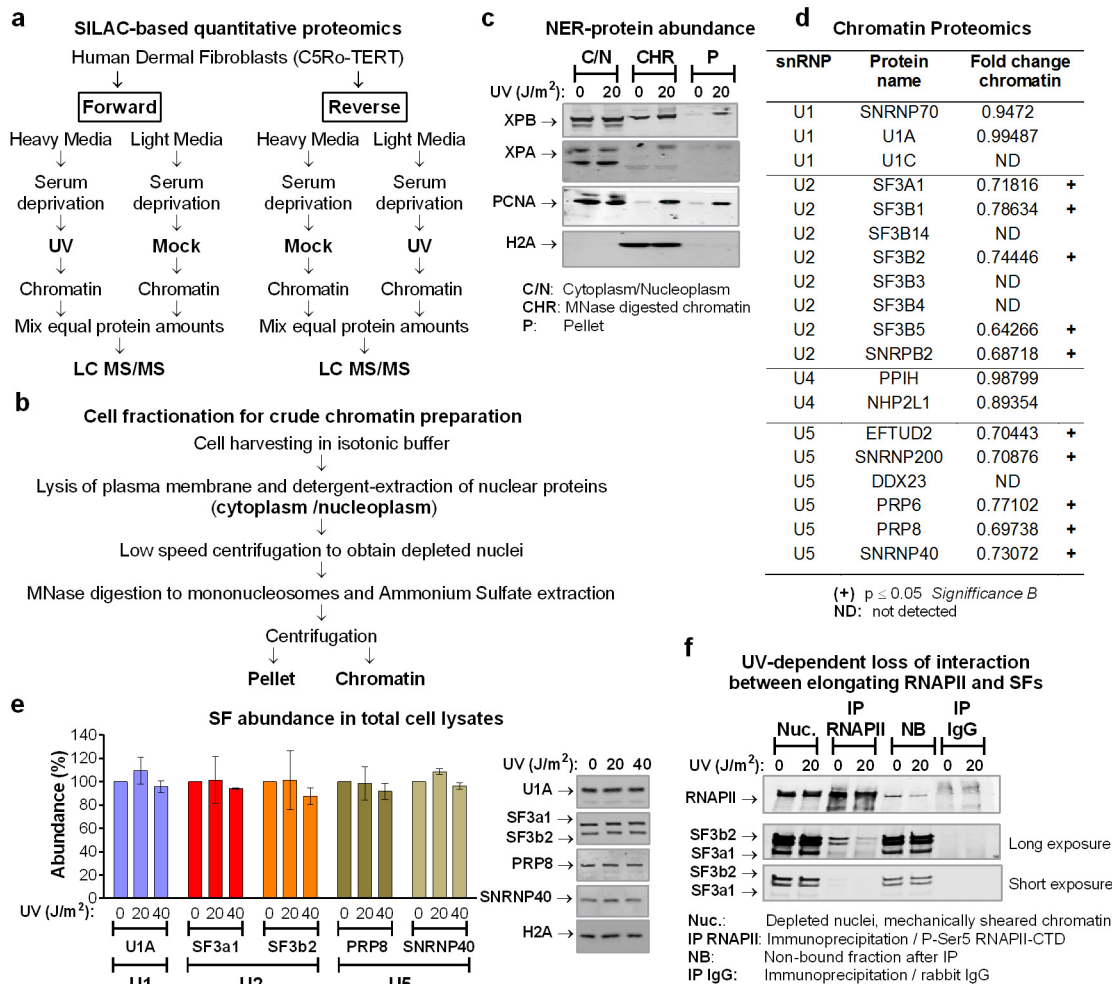
**Paired-end RNA-seq and data analysis.** RPE-1 cells were grown to confluence and serum-deprived for 72 h to ensure quiescence. Cells were then mock-treated or UV-irradiated with 20 J m<sup>–2</sup> UVC (245 nm), in the presence or absence of 10 µM of the ATM inhibitor. Each treatment was performed in duplicate plates that were used as biological replicates. RNA was isolated 6 h post-irradiation with the RNeasy kit (QIAGEN) and PolyA RNA was isolated using the Dynabeads mRNA purification kit (Invitrogen). Sample integrity was verified by the Agilent 2100 Bioanalyzer (Agilent Technologies). For all samples the Bioanalyzer RNA integrity scores (RIN) were 9.3–10, indicating excellent RNA quality. For each sample a cDNA library was prepared and validated using the Illumina TruSeq RNA sample preparation kit v2 according to the manufacturer's instructions. In brief, equal amounts (200 ng) of poly(A)-RNA were chemically fragmented, cDNA was generated using random hexamers as primers and adapters were ligated. RNA fragmentation efficiency and similarity between samples was confirmed by the Bioanalyzer after adaptor ligation (average fragment sizes were 317–344 bp). Following PCR amplification RNA-seq was performed according to the Illumina TruSeq v3 protocol on the HiSeq2500 platform, generating paired-end, 100-bp reads (9 × 10<sup>7</sup> reads per sample). Raw reads were aligned against the human genome assembly (hg19) using TopHat<sup>44</sup>. Uniquely mapped reads were used for the identification of alternative splicing events using multi-variate analysis of transcript splicing (MATS) as previously described<sup>45</sup>. Alternative splicing events that were significantly increased ( $P < 0.05$ ,  $n = 2$ ) by UV irradiation (untreated versus UV-irradiated cells) by a minimum of 10% difference were considered to be UV-induced. Each individual alternative splicing event identified in the first analysis to be UV-induced that also decreased by a minimum of 10% in UV-irradiated cells in absence of ATM activity (UV-irradiated cells versus ATM-inhibitor-treated/UV-irradiated cells) was considered to be (partly) dependent on ATM activity.

**Statistical analysis.** All data presented were reproduced in at least three independent experiments. Statistical analysis was performed using the PRIZM GraphPad software unless otherwise stated. Significance of differences was evaluated with either Student's *t*-test, when only two groups were compared, or

one-way ANOVA for more than two groups. No statistical methods were used to predetermine sample size. One-way ANOVA was followed by post hoc analysis either by Dunnett's test (for comparison of experimental conditions to control) or Bonferroni's test (comparison between groups). \* $P < 0.05$ , \*\* $P < 0.01$  and \*\*\* $P < 0.001$ . Proteomic data statistical analysis was performed with Perseus (1.5.0.30)<sup>17</sup>. Significance B was calculated by estimating the variance of the distribution of all protein ratios, taking into account the dependency of the distribution on the summed protein intensity<sup>17</sup>. Peptides with a significance B value  $P \leq 0.05$  in either the forward or reverse experiment were considered significant and indicated by (+) in Extended Data Fig. 1d and Supplementary Table 1. Significant alternative splicing events were identified by MATS<sup>45</sup>; only UV-triggered events with  $P < 0.05$  were used for further analysis.

39. Schwertman, P. *et al.* UV-sensitive syndrome protein UVSSA recruits USP7 to regulate transcription-coupled repair. *Nature Genet.* **44**, 598–602 (2012).
40. Nakazawa, Y., Yamashita, S., Lehmann, A. R. & Ogi, T. A semi-automated non-radioactive system for measuring recovery of RNA synthesis and unscheduled DNA synthesis using ethynyluracil derivatives. *DNA Repair (Amst.)* **9**, 506–516 (2010).
41. Houtsmuller, A. B. & Vermeulen, W. Macromolecular dynamics in living cell nuclei revealed by fluorescence redistribution after photobleaching. *Histochem. Cell Biol.* **115**, 13–21 (2001).
42. Wuarin, J. & Schibler, U. Physical isolation of nascent RNA chains transcribed by RNA polymerase II: evidence for cotranscriptional splicing. *Mol. Cell. Biol.* **14**, 7219–7225 (1994).
43. Galiveti, C. R., Rozhdestvensky, T. S., Brosius, J., Lehrach, H. & Konthur, Z. Application of housekeeping npcRNAs for quantitative expression analysis of human transcriptome by real-time PCR. *RNA* **16**, 450–461 (2010).
44. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
45. Shen, S. *et al.* MATS: a Bayesian framework for flexible detection of differential alternative splicing from RNA-Seq data. *Nucleic Acids Res.* **40**, e61 (2012).

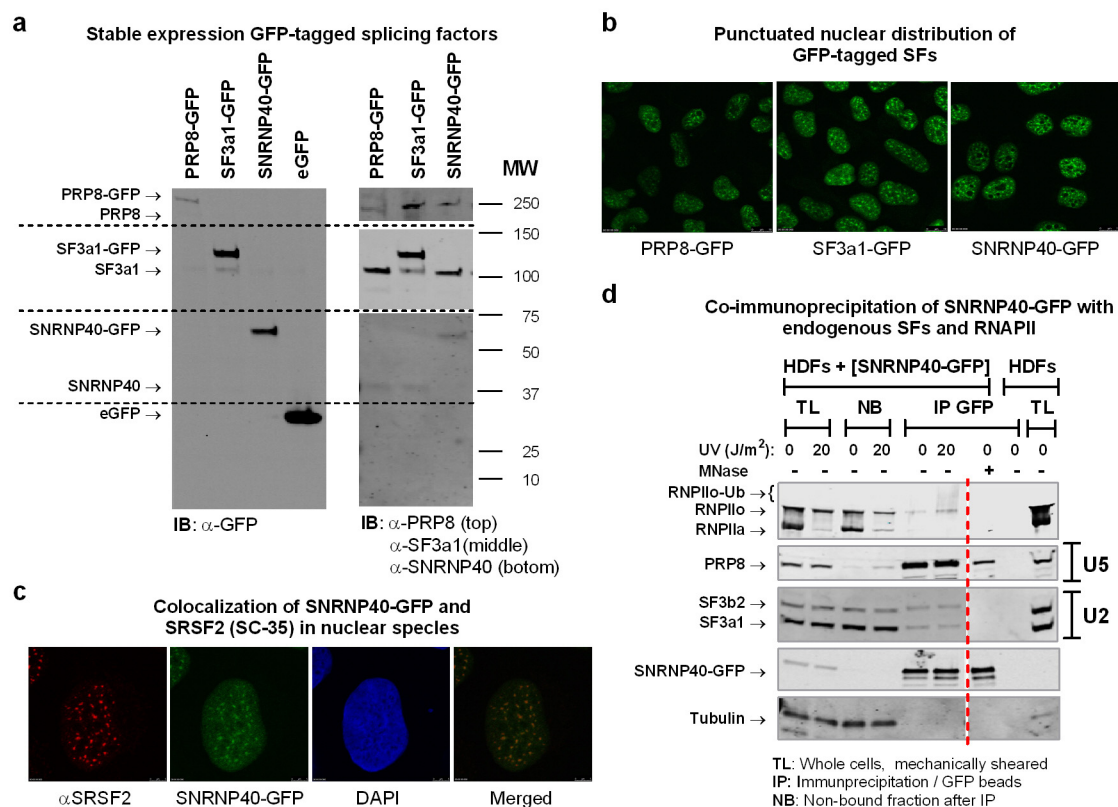




# Extended Data Figure 1 | Chromatin association of splicing factors.

**a**, Schematic overview of the proteomic experiments for the identification of proteins that display UV-dependant chromatin association. **b**, Schematic outline of cell fractionation. **c**, Validation of the chromatin-isolation protocol for NER proteins that are recruited to chromatin in response to DNA damage. Mock-treated or UV-irradiated quiescent HDFs ( $20 \text{ J m}^{-2}$ , 1 h post-irradiation) were fractionated as outlined in **b**. Equal protein amounts from each fraction were analysed by immunoblotting using antibodies against the indicated NER proteins. Abundance of H2A is shown as a control for chromatin-isolation efficiency. **d**, UV-triggered changes in chromatin association of core splicing factors, identified by quantitative SILAC proteomics. Proteomic experiments were performed with HDFs as outlined in **a**. The table lists representative examples of splicing factors that participate in distinct snRNP complexes and their chromatin association in response to UV irradiation ( $20 \text{ J m}^{-2}$ , 1 h). U2 and U5 snRNP splicing factors show significantly reduced chromatin association ( $P \leq 0.05$ , significance B<sup>17</sup>) and are

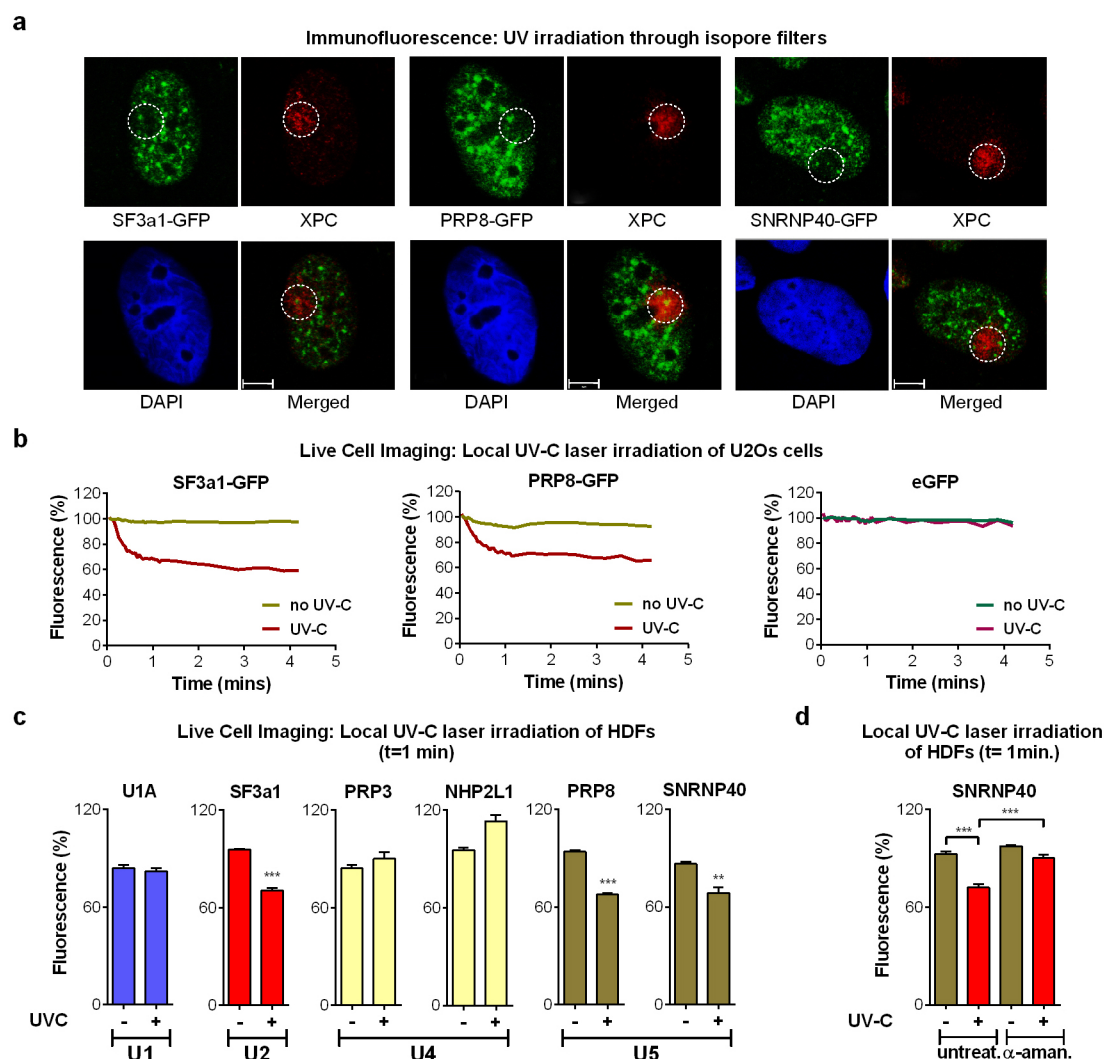
indicated with a cross. Significance B was calculated by estimating the variance of the distribution of all protein ratios, taking into account the dependency of the distribution on the summed protein intensity<sup>17</sup>. ND, not detected. **e**, Abundance of splicing factors in total cell lysates. Total lysates were prepared from U2OS cells that were mock-treated or UV-irradiated ( $20 \text{ J m}^{-2}$ , 1 h post-irradiation) and splicing factor abundance was assayed by immunoblotting. Abundance of H2A is shown as a loading control. Right, immunoblots; left, quantification of splicing factor signal intensities normalized to H2A ( $n = 3$ , mean  $\pm$  s.d., one-way ANOVA/Bonferroni). **f**, UV-dependent interaction of splicing proteins with elongating RNAPII. Quiescent HDFs were prepared as outlined in **b** except that, instead of MNase digestion, chromatin was mechanically sheared. Elongating RNAPII was immunoprecipitated with an antibody that recognizes specifically the Ser2-phosphorylated RNAPII C-terminal domain (CTD) and its interaction with the U2 snRNP splicing factors SF3a1 and SF3b2 was assayed by immunoblotting.



**Extended Data Figure 2 | Validation of HDFs stably expressing GFP-tagged splicing factors.** **a**, Whole-cell lysates from HDFs stably expressing eGFP tagged PRP8, SF3a1, SNRNP40 or free eGFP, were analysed by immunoblotting using antibodies against GFP (left) or against PRP8, SF3a1 and SNRNP40 (right). Ectopically produced proteins were expressed at near or below endogenous levels. **b**, Fluorescent microscopy images of GFP-tagged splicing factors showing the expected punctuated nuclear distribution. Images were obtained at 40 $\times$  magnification. **c**, Localization of SNRNP40-GFP in nuclear speckles which were visualized by immunofluorescence detection of the speckle marker SRSF2/SC35. Images were obtained at 63 $\times$  magnification. **d**, Interaction of SNRNP40-GFP with endogenous splicing factors and elongating RNAPII. Quiescent SNRNP40-GFP-expressing HDFs were mock-treated or irradiated with 20  $J\ m^{-2}$  UVC. After a 3-h recovery period, cells

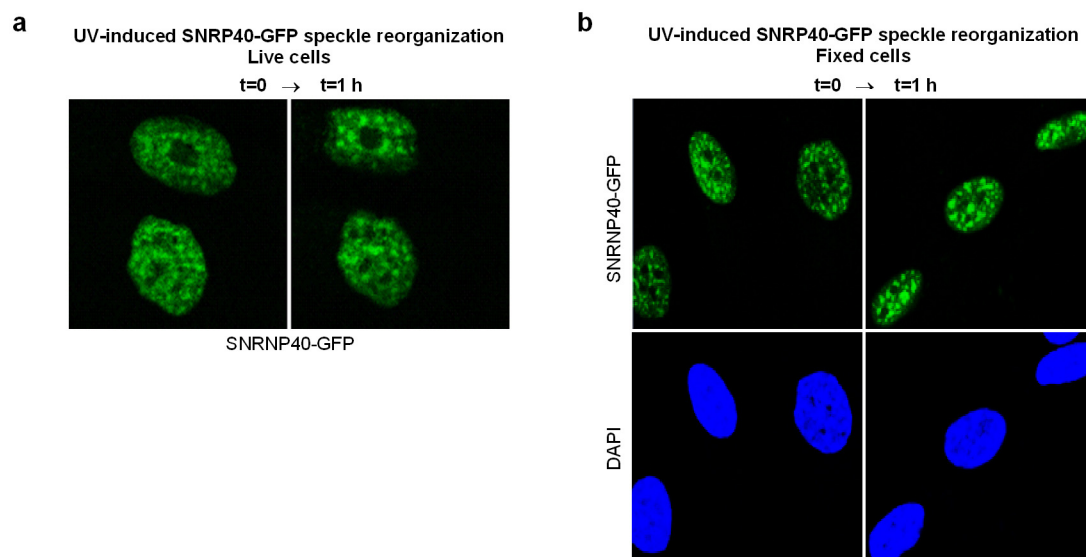
were lysed under native conditions and chromatin was sheared by mechanical force. SNRNP40-GFP was immunoprecipitated from whole-cell lysates using GFP-Trap agarose beads, and its association with endogenous splicing factors and the large subunit of RNAPII was assayed by immunoblotting. Non-transfected cells are shown as a negative control. SNRNP40-GFP interacts with U2 and U5 snRNP components, suggesting that the GFP tag does not interfere with complex formation. Interaction of SNRNP40 with its U5 snRNP partner PRP8 is partially maintained even after MNase digestion, consistent with its presence in U4/U6.U5 tri-snRNP complexes. Participation of SNRNP40-GFP in co-transcriptional splicing complexes is confirmed by co-immunoprecipitation of the active (hyperphosphorylated RNAPII $\alpha$ ) large subunit of RNAPII.





**Extended Data Figure 3 | Displacement of mature spliceosomes from subnuclear sites of UV-inflicted DNA damage.** **a**, U2OS cells stably expressing GFP-tagged splicing factors were UV-irradiated ( $60 \text{ J m}^{-2}$ ) through isopore membranes resulting in DNA lesion formation in small subnuclear areas. DNA damage sites (circled) were visualized by immunofluorescence using an antibody against the NER recognition factor XPC. Scale bar,  $5 \mu\text{m}$ . **b**, SF3a1-GFP and PRP8-GFP depletion from UVC laser microbeam irradiation sites. Quantification of 20 cells from two independent experiments. eGFP localization at sites of DNA damage is used to demonstrate that depletion of eGFP-tagged splicing factors is not caused by photobleaching. **c**, UVC laser microbeam irradiation results in preferential displacement of U2- and U5-associated splicing factors from DNA damage sites. Quiescent HDFs were irradiated in a  $\sim 1\text{-}\mu\text{m}$ -diameter nuclear area via a UVC laser. GFP signal

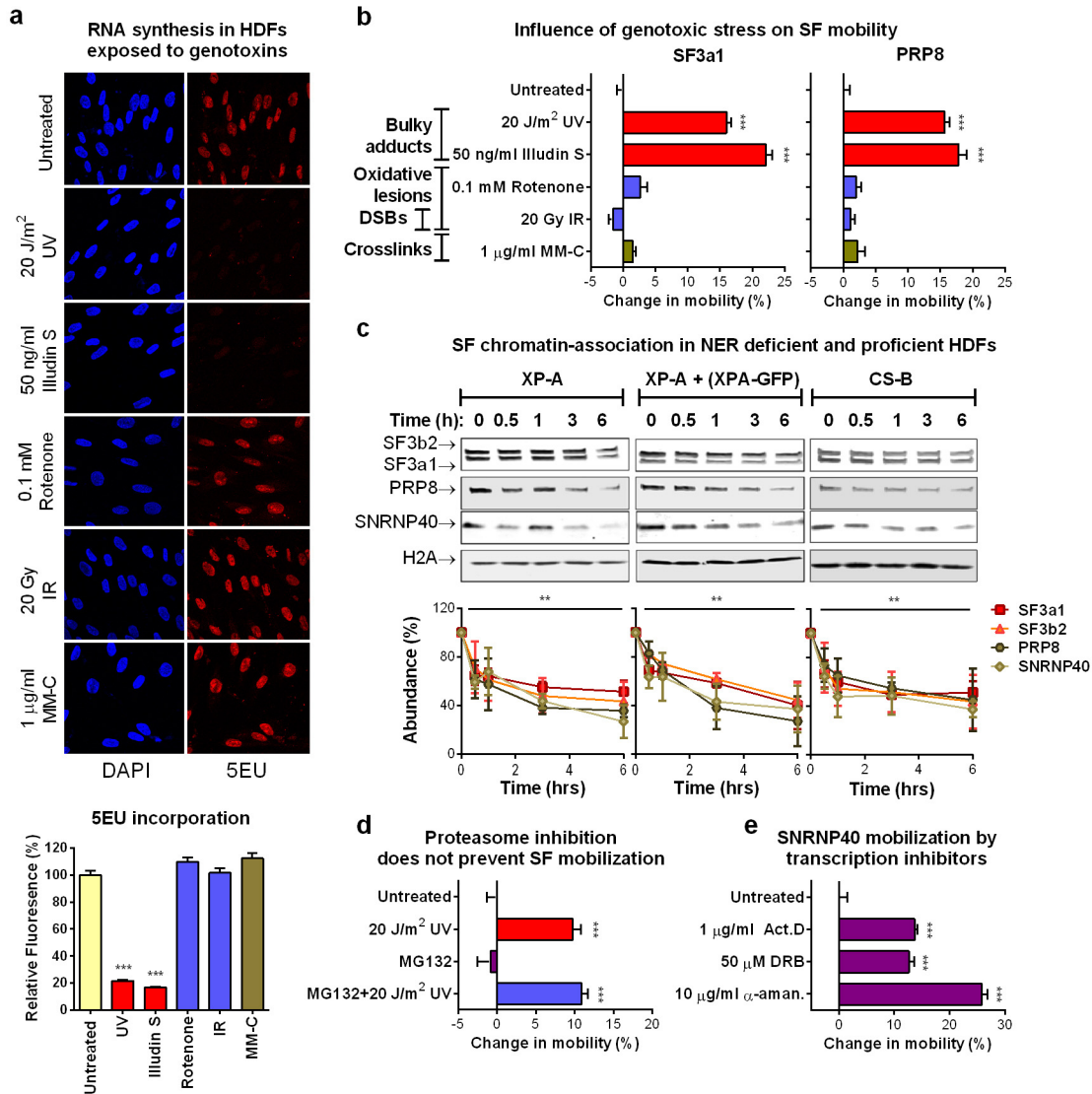
intensity, reflecting the abundance of GFP-tagged U1, U2, U4 and U5 snRNP components at UVC DNA damage sites, was quantified in the irradiated and in a non-irradiated nuclear area (undamaged control). Plotted is the fluorescence signal intensity expressed as a percentage of that before irradiation, at the 1-min time point. Cells expressing free eGFP were used as negative control. Representative from three independent experiment ( $n = 12$ , mean  $\pm$  s.e.m., paired  $t$ -test). **d**, Depletion of splicing factors from UVC laser irradiation sites depends on active transcription. Transcription initiation was inhibited in quiescent HDFs by prolonged  $\alpha$ -amanitin treatment ( $10 \mu\text{M}$ ,  $\geq 24 \text{ h}$ ) before subnuclear UVC laser irradiation. Plotted is the SNRNP40-GFP abundance in irradiated and non-irradiated nuclear areas at 1-min post-irradiation. Representative from three independent experiments ( $n = 12$ , mean  $\pm$  s.e.m., one-way ANOVA/Bonferroni).



**Extended Data Figure 4 | SNRNP40 reorganization and speckle enlargement in response to UV irradiation.** Representative microscopic images showing SNRNP40-GFP distribution in quiescent HDFs before, and

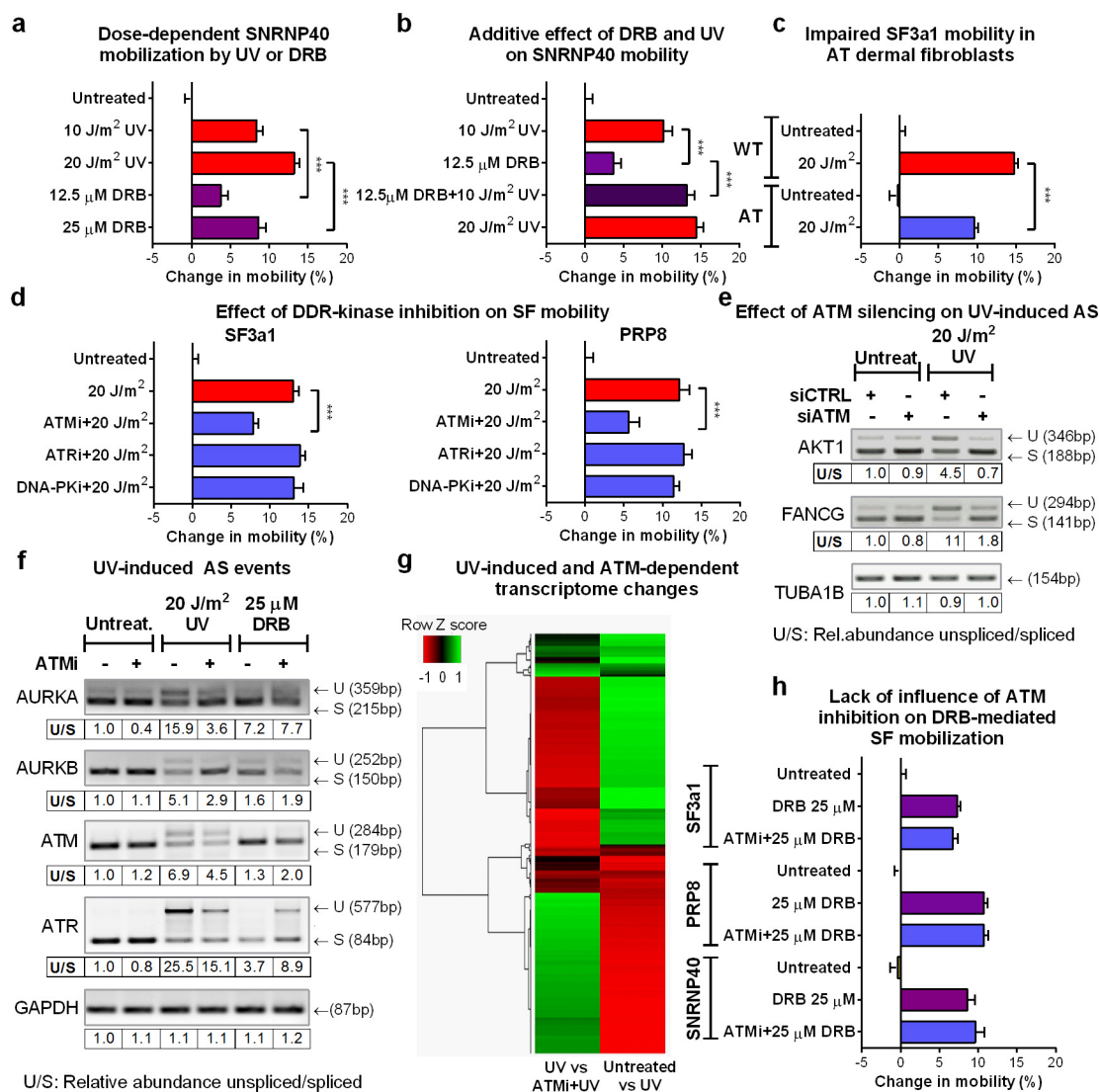
1 h post UVC irradiation with  $20 \text{ J m}^{-2}$ . **a**, Live cells. **b**, Fixed cells. Images were obtained at  $63\times$  magnification.





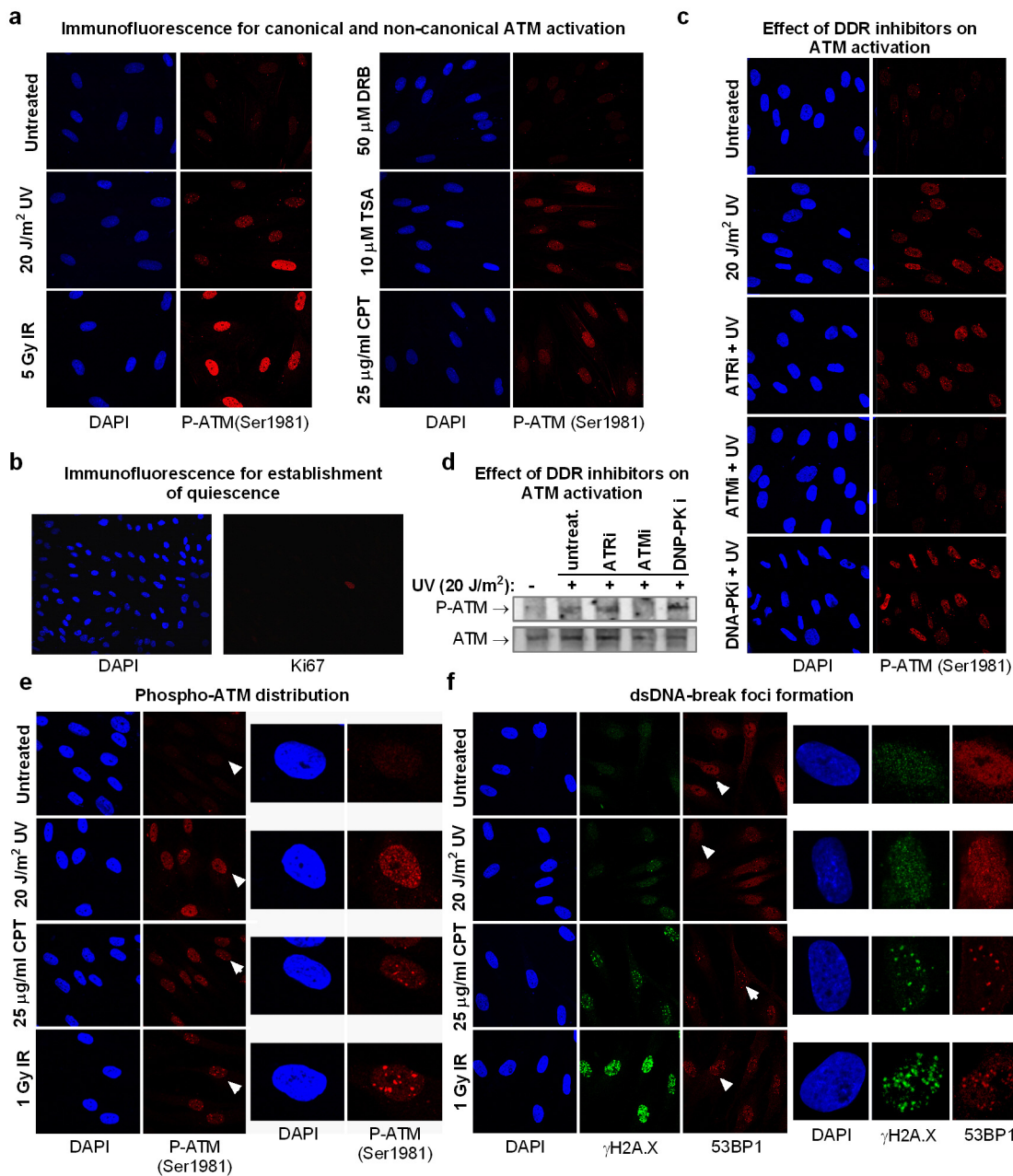
**Extended Data Figure 5 | Transcription stalling mobilizes spliceosomes independent from NER complex assembly and proteasome activity.** **a**, RNA synthesis is inhibited preferentially by genotoxins that inflict bulky DNA lesions. Influence of genotoxins on RNA synthesis of quiescent HDFs was measured by 5EU pulse labelling combined with click chemistry. Top, representative images; bottom, quantification of fluorescence intensity ( $n = 150$ , mean  $\pm$  s.e.m., one-way ANOVA/Bonferroni). Images were obtained at 40 $\times$  magnification. **b**, Mobilization of U2 and U5 snRNPs by genotoxins inflicting transcription-blocking DNA lesions. Mobilization of GFP-tagged SF3a1 (left) and PRP8 (right) assayed by FRAP in quiescent HDFs exposed to different types of genotoxins ( $n = 30$ , mean  $\pm$  s.e.m., one-way ANOVA/Dunnett's). IR, ionizing radiation. **c**, Chromatin displacement of mature spliceosomes is not TC-NER-dependent. Left, chromatin abundance of U2 and

U5 snRNP splicing factors was assayed by immunoblotting in XPA-deficient (left), XPA-GFP-corrected (middle) and CSB-deficient (right) HDFs. Cells were UV-irradiated ( $20 \text{ J m}^{-2}$ ) and chromatin was isolated at the indicated times. Top, immunoblots; bottom, quantification of splicing factor signal intensities normalized to H2A ( $n = 3$ , mean  $\pm$  s.d., one-way ANOVA/Bonferroni). **d**, Proteasome activity is not required for UV-damage-induced spliceosome mobilization. Mobilization of SNRNP40-GFP assayed by FRAP in quiescent HDFs exposed to UV radiation in the presence or absence of the proteasome inhibitor MG132 ( $50 \mu\text{M}$ ) ( $n = 30$ , mean  $\pm$  s.e.m.,  $t$ -test). **e**, SNRNP40-GFP mobilization by transcription inhibition. FRAP of SNRNP40-GFP in quiescent HDFs after inhibition of transcription initiation ( $10 \mu\text{g ml}^{-1}$   $\alpha$ -amanitin, 24 h) or elongation ( $1 \mu\text{g ml}^{-1}$  actinomycin D or  $50 \mu\text{M}$  DRB, 1 h) ( $n = 30$ , mean  $\pm$  s.e.m., one-way ANOVA/Dunnett's).



**Extended Data Figure 6 | ATM-dependency of UV induced spliceosome mobilization, alternative splicing and gene expression changes.** **a**, UV irradiation and DRB-dependent mobilization of SNRNP40. Quiescent HDFs expressing SNRNP40-GFP were UV-irradiated or DRB treated with doses that inhibit transcription to similar levels. Splicing factor mobility was assayed by FRAP. **b**, Additive effect of combined UV and DRB treatments. FRAP of SNRNP40-GFP in quiescent HDFs treated with DRB, UV, or a combination of both, each at a dose that inhibits RNA synthesis by ~50%. **c**, Impaired UV-dependent SF3a1 mobilization in cells lacking ATM activity. SF3a1-GFP mobilization was measured by FRAP in quiescent HDFs derived from an ataxia telangiectasia (AT) patient or a healthy donor. **d**, ATM-dependent spliceosome mobilization. Quiescent HDFs were treated with 10  $\mu$ M ATM (KU55933), ATR (VE821) or DNA-PK (NU7441) inhibitors before irradiation. GFP-tagged SF3a1 or PRP8 mobility was assayed by FRAP. ATM, but not ATR or DNA-PK inhibition partially prevented the UV-induced splicing factor mobilization. **a–d**,  $n = 25$ , mean  $\pm$  s.e.m., one-way ANOVA/Bonferroni. **e**, Reduced UV-induced intron retention in response to ATM silencing. Intron inclusion in

retina pigment epithelium (RPE) cells transfected either with control or ATM-silencing siRNAs and subsequently mock-treated or UV-irradiated ( $20 \text{ J m}^{-2}$ , 6 h) was assayed by RT-PCR. **f**, ATM-dependent changes in intron retention. Intron inclusion was assayed by RT-PCR in untreated, UV-irradiated and DRB-treated quiescent cells in the presence or absence of 10  $\mu$ M ATM inhibitor. **g**, Heat map of UV-triggered and ATM-dependent transcriptome changes. Quiescent cells were mock-treated or UV-irradiated in the presence or absence of the ATM inhibitor. Transcriptome profiles were generated by RNA-seq. Differentially expressed genes between untreated and UV-irradiated cells ( $P < 0.05$ ) and UV-irradiated cells in the presence or absence of the ATM inhibitor ( $P < 0.05$ ), were clustered in a heat map using Pearson correlation.  $n = 1,676$  differentially expressed transcripts. The observed correlation indicates that UV-inducible transcriptome changes can be, in part, prevented by ATM inhibition. **h**, Lack of influence of ATM inhibition on DRB-mediated SF mobilization. Splicing factor mobility was measured by FRAP in untreated or DRB-treated HDFs in the presence or absence of 10  $\mu$ M ATM inhibitor ( $n = 30$ , mean  $\pm$  s.e.m., one-way ANOVA/Bonferroni).

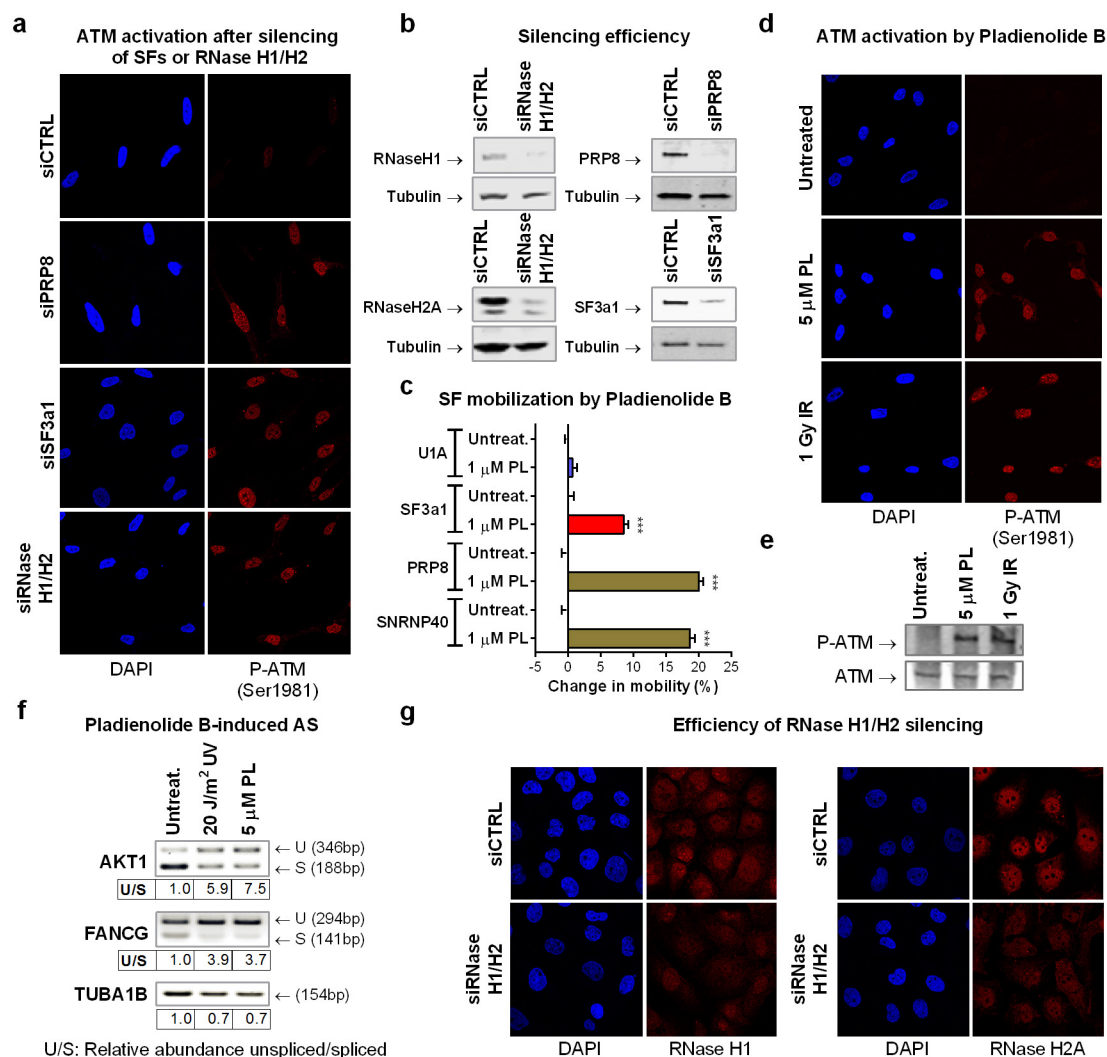


#### Extended Data Figure 7 | Canonical and non-canonical ATM activation.

**a**, ATM autophosphorylation (Ser1981) was assayed in quiescent HDFs 1 h after the indicated treatments. In non-replicating cells UV and trichostatin A (TSA) activate ATM via non-canonical pathways. Transcription inhibition by DRB has no influence on ATM activity. **b**, The quiescent status of serum-deprived HDFs was verified by immunodetection of the cell cycle marker Ki67, which is not expressed by quiescent (G<sub>0</sub>) cells. **c**, Immunofluorescence detection of active ATM in quiescent HDFs treated with DDR kinase inhibitors. **d**, Immunoblotting analysis of nuclear extracts derived from quiescent HDFs treated as in **c** using a phospho-specific ATM (Ser1981) antibody (top) and an antibody recognizing ATM (bottom). **e**, Differences in autophosphorylated-ATM distribution in quiescent HDFs treated with various ATM activators. Left,

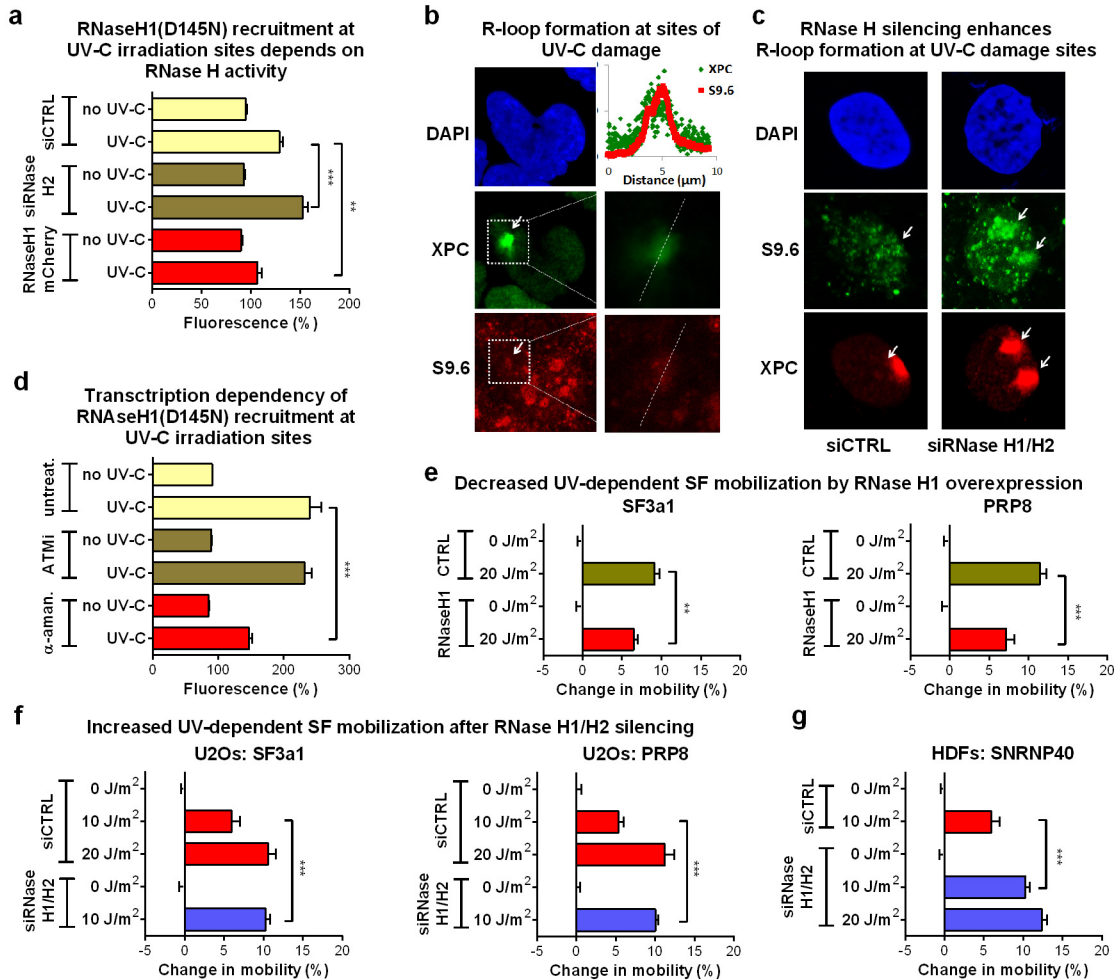
multiple cells; right, single-cell magnification illustrating pan-nuclear localization of phosphorylated ATM after UV irradiation and focal accumulation after CPT or ionizing radiation treatments. Magnified cells are indicated by arrows (left panel). **f**, Differences in amounts of DNA damage-foci formation indicative of DSBs, in response to CPT, UV and ionizing radiation. Quiescent HDFs were pre-treated with the ATR inhibitor (10  $\mu$ M, 1 h) and subsequently exposed to the indicated genotoxins. DSB foci were visualized by immunofluorescence using antibodies against  $\gamma$ H2AX and p53BP1. Left, multiple cells; right, single-cell magnification; images were obtained at 40 $\times$  and 63 $\times$  magnification, respectively. Magnified cells are indicated by arrows in the left panel.





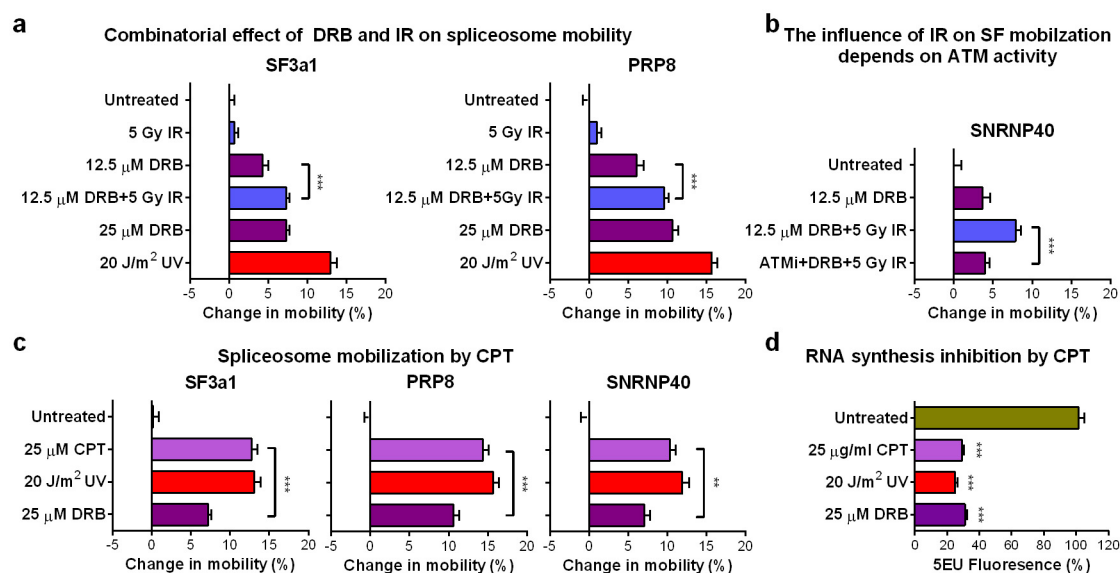
**Extended Data Figure 8 | ATM activation by interference with spliceosome assembly or RNaseH1/H2A silencing.** **a**, ATM autophosphorylation was assayed by immunofluorescence in HDFs after silencing of SF3a1, PRP8 or combined silencing of RNaseH1 and RNaseH2A. **b**, Immunoblotting analysis of silenced proteins in total cell lysates. Tubulin is shown as a loading control. **c**, Splicing factor mobilization by the spliceosome inhibitor pladienolide B (PL) was assayed by FRAP in quiescent HDFs. Consistent with its function in interfering with spliceosome maturation following pre-spliceosome assembly, cell treatment with pladienolide B resulted in extensive mobilization of U5 snRNP factors (PRP8 and SNRNP40), partial mobilization of the U2 snRNP SF3a1, and had no influence on the U1 snRNP factor U1A ( $n = 30$ ,

mean  $\pm$  s.e.m., one-way ANOVA/Bonferroni). **d**, **e**, ATM activation by Pladienolide B. Quiescent HDFs were either treated with 5  $\mu$ M pladienolide B or exposed to 1 Gy ionizing radiation (IR) and autophosphorylated ATM was detected by immunofluorescence (**d**) or immunoblotting (**e**). **f**, Effect of pladienolide treatment on intron retention. RNA isolated from mock-treated, UV-irradiated or pladienolide B-treated RPE cells. Intron retention, as assayed by RT-PCR on transcripts of the indicated genes, shows that pladienolide B influences splicing to the same extent as UV irradiation. U/S, ratio of relative abundance of unspliced (U) to spliced (S) introns. **g**, Efficiency of RNaseH1 and H2A silencing at the single-cell level, assayed by immunofluorescence. Images were obtained at 40 $\times$  magnification.



**Extended Data Figure 9 | UV-induced R-loop formation enhances spliceosome mobilization.** **a**, Recruitment of RNaseH1(D145N)-GFP at local DNA damage sites depends on endogenous levels of RNaseH activity. DNA damage was inflicted via a UVC laser in  $\sim 1\text{-}\mu\text{m}$ -diameter subnuclear areas of cells after silencing of RNaseH2A or overexpression of RNaseH1-mCherry. Recruitment of RNaseH1(D145N)-GFP at the irradiated sites was monitored by live-cell imaging. Plotted is the fluorescence intensity of RNaseH1(D145N)-GFP at 1 min post-irradiation, at the irradiated and in a non-irradiated nuclear area. Representative from three independent experiments ( $n = 10$ , mean  $\pm$  s.e.m., one-way ANOVA/Bonferroni). **b, c**, R-loop formation at sites of local UVC laser irradiation. Immunofluorescence detection of R-loops using the DNA-RNA hybrid-specific S9.6 antibody. Sites of irradiation are visualized by XPC immunodetection. **b**, Dashed boxes indicate the magnified areas shown in the right panels. The dashed lines indicate the line-scan track used to quantify fluorescence intensity of S9.6 and anti-XPC (shown in the graph). **c**, Specificity of the antibody was confirmed by its increased sensitivity after RNaseH2A silencing and its ability to detect R-loops when suboptimal doses of

UVC irradiation were applied. **d**, RNaseH1 accumulation at local DNA damage sites depends on active transcription but not ATM activity. Transcription initiation was inhibited in quiescent HDFs by  $\alpha$ -amanitin ( $10\text{ }\mu\text{g ml}^{-1}$ , 24 h) before local UVC laser irradiation. Plotted is the fluorescence intensity at 1 min post-irradiation of RNaseH1(D145N)-GFP at the irradiated area and in a non-irradiated nuclear area for untreated, ATM-inhibitor- and  $\alpha$ -amanitin-treated cells. Representative from three experiments ( $n = 10$ , mean  $\pm$  s.e.m., one-way ANOVA/Bonferroni). **e**, RNaseH1 overexpression inhibits the UV-dependent spliceosome mobilization. FRAP of U2OS cells stably expressing GFP-tagged SF3a1 and PRP8 and transiently transfected with RNaseH1-mCherry. **f**, RNaseH1 and H2A silencing potentiates the UV-dependent spliceosome mobilization. RNaseH1 and H2 were silenced in U2OS cells expressing SF3a1-GFP or PRP8-GFP and splicing factor mobility was assayed by FRAP. **g**, FRAP of SNRNP40-GFP in quiescent HDFs after RNaseH1/H2 silencing. **e-g**,  $n = 30$ , mean  $\pm$  s.e.m., one-way ANOVA/Bonferroni.



### Extended Data Figure 10 | Combined transcription inhibition and ATM activation results in extensive mobilization of mature spliceosomes.

**a**, Combinatorial effect of DRB and ionizing radiation on spliceosome mobilization. Quiescent HDFs were exposed to ionizing radiation in the presence or absence of DRB, and SF3a1-GFP and PRP8-GFP mobility was assayed by FRAP. **b**, The ionizing-radiation-mediated increase of DRB-dependent spliceosome mobilization depends on ATM activity. FRAP of GFP-tagged SNRNP40 in quiescent HDFs treated with DRB and/or ionizing radiation in the presence or absence of an ATM inhibitor. **c**, Spliceosome

mobilization by CPT. Quiescent HDFs were treated with 25  $\mu$ g ml<sup>-1</sup> CPT, 25  $\mu$ M DRB and 20 J m<sup>-2</sup> UV at doses that inhibit transcription to approximately 30% and their influence on SF3a1, PRP8 and SNRNP40 mobilization was measured by FRAP. Mobilization of GFP-tagged SF3a1, PRP8 and SNRNP40 in quiescent HDFs was measured by FRAP. **a-c**,  $n = 30$ , mean  $\pm$  s.e.m., one-way ANOVA/Bonferroni. **d**, Inhibition of RNA synthesis by the treatments shown in **c** was assayed in quiescent HDFs by 5EU incorporation and click chemistry ( $n = 150$ , mean  $\pm$  s.e.m., one-way ANOVA/Dunnett's).



# Self-similar energetics in large clusters of galaxies

Francesco Miniati<sup>1</sup> & Andrey Beresnyak<sup>2</sup>

Massive galaxy clusters are filled with a hot, turbulent and magnetized intra-cluster medium. Still forming under the action of gravitational instability, they grow in mass by accretion of supersonic flows. These flows partially dissipate into heat through a complex network of large-scale shocks<sup>1</sup>, while residual transonic (near-sonic) flows create giant turbulent eddies and cascades<sup>2,3</sup>. Turbulence heats the intra-cluster medium<sup>4</sup> and also amplifies magnetic energy by way of dynamo action<sup>5–8</sup>. However, the pattern regulating the transformation of gravitational energy into kinetic, thermal, turbulent and magnetic energies remains unknown. Here we report that the energy components of the intra-cluster medium are ordered according to a permanent hierarchy, in which the ratio of thermal to turbulent to magnetic energy densities remains virtually unaltered throughout the cluster's history, despite evolution of each individual component and the drive towards equipartition of the turbulent dynamo. This result revolves around the approximately constant efficiency of turbulence generation from the gravitational energy that is freed during mass accretion, revealed by our computational model of cosmological structure formation<sup>3,9</sup>. The permanent character of this hierarchy reflects yet another type of self-similarity in cosmology<sup>10–13</sup>, while its structure, consistent with current data<sup>14–18</sup>, encodes information about the efficiency of turbulent heating and dynamo action.

The computational model we use captures the turbulent motions through a multi-scale technique that employs six nested grids covering progressively larger volumes with correspondingly coarser resolution elements<sup>3,19</sup>. The finest grid resolves the virial volume of the galaxy cluster (GC) with more than a billion uniformly-sized resolution elements, and provides the necessary dynamic range to resolve the intra-cluster medium (ICM) turbulent cascade. The largest grid covers the chosen cosmological volume of 340 co-moving Mpc on a side ('co-moving' indicates that the system is taking part in the expansion of the universe; 1 Mpc  $\approx 3 \times 10^6$  light years). The intermediate grids allow us to account, with adequate accuracy, for the tidal forces produced by the matter distribution outside the GC volume. The calculation starts with three grids and adds progressively finer grids as the Lagrangian volume of the GC shrinks under self-gravity. All six grids are in place at a time corresponding to 8 Gyr after the Big Bang. At the current time, 13.8 Gyr after the Big Bang, the simulated GC has a total virial mass of  $1.3 \times 10^{15}$  solar masses ( $M_\odot$ ).

Figure 1 shows a snapshot of the simulation, illustrating the cosmological context and the highly-turbulent conditions of the flow inside the GC volume. The fine resolution across the GC volume allows us to accurately measure the time-dependent statistical properties of structure-formation-driven ICM turbulence including, in particular, the dissipation rate,  $\epsilon_{\text{turb}}$ , the outer scale,  $L$ , and the dispersion of the velocity increment on that scale,  $\langle(\delta u_L)^2\rangle^{1/2}$  (Methods). In the following, we restrict our analysis to a region within  $R_{\text{vir}}/3 \approx 1$  co-moving Mpc, which is most relevant for comparison with observations, where the virial radius  $R_{\text{vir}}$  characterizes a volume that has nominally reached dynamical equilibrium (Methods).

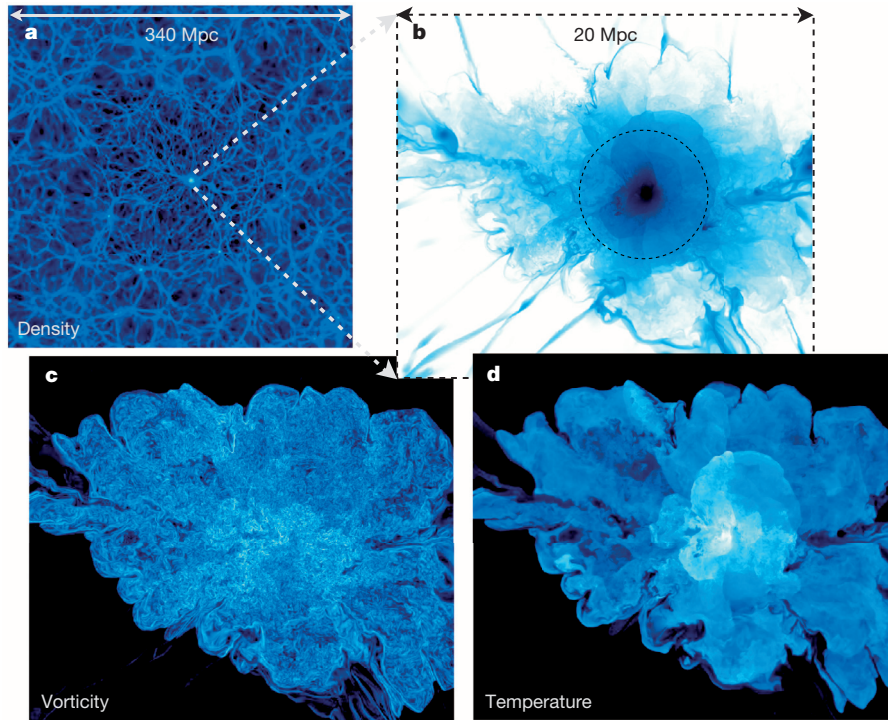
Hydrodynamic turbulence is dominated by the solenoidal component, which accounts for 60–90% of the total kinetic energy<sup>9,20</sup>. Detailed analysis shows that this component remains statistically homogenous and isotropic thus resembling Kolmogorov's cascade, despite the presence of considerable structure in the ICM<sup>9</sup>. The dissipation of incompressible turbulence contributes to ICM heating<sup>4</sup> along with shocks and adiabatic compression, and to the growth of magnetic energy by way of small-scale dynamo action<sup>5,8</sup> (see Extended Data Fig. 1 for cascade details). The turbulent dissipation rate associated with the solenoidal component is estimated from the numerical simulation data. Because ICM turbulence is driven by various complex hydrodynamic mechanisms ultimately powered by the unsteady mass accretion process<sup>9</sup>, the dissipation rate is highly changeable with time and exhibits non-monotonic variations of more than one order of magnitude<sup>20</sup> (Fig. 2c).

However, alongside much complexity, turbulent dissipation appears to also exhibit simplicity of behaviour. This is shown in Fig. 2a, which illustrates the temporal evolution of the fraction of thermal energy originating from turbulent dissipation,  $\eta_{\text{turb}}$ . In contrast to  $\epsilon_{\text{turb}}$ , this quantity remains remarkably constant during the GC lifetime,  $\eta_{\text{turb}} \approx 0.3$ – $0.4$ , indicating that the efficiency of turbulence generation out of gravitational energy freed by mass accretion is approximately constant. In addition, Fig. 2b shows that the turbulence velocity dispersion at the outer scale normalized to the ICM sound speed ( $c_s$ ), that is, the turbulence Mach number  $\mathcal{M}_{\text{turb}} = \langle(\delta u_L)^2\rangle^{1/2}/c_s$ , also remains almost constant with time. This is certainly related to the fact that the GC remains dynamically young, that is, it continues to accrete mass and grow in size, but also shows that in the ICM the evolution of the turbulent kinetic energy and the thermal energy are closely related, consistent with Fig. 2a. The value of  $\mathcal{M}_{\text{turb}}$  can be understood as follows. If the generation of the bulk of the thermal energy,  $E_{\text{th}}$ , is dominated by the last  $\alpha$  eddy turnover times,  $\tau_L = L/\delta u_L$ , then  $E_{\text{th}} \approx \eta_{\text{turb}}^{-1} \int \rho \epsilon_{\text{turb}} dt \approx [\alpha/(3^{3/2} \eta_{\text{turb}})] \rho \langle(\delta u_L)^2\rangle$ , where  $\rho$  is the mass density and we have used the known relation  $\epsilon_{\text{turb}} = (2/3C)^{3/2} \langle(\delta u_L)^2\rangle^{3/2}/L$  with the coefficient of the correlation function  $C \approx 2$  (ref. 21). It is straightforward to then see that the Mach number  $\mathcal{M}_{\text{turb}} \approx (\sqrt{3}/\alpha)^{1/2} (\eta_{\text{turb}}/0.37)^{1/2}$  which, for  $\alpha = 1.5$ – $3$ , ranges between 0.8 and 1.2, confirming the result in Fig. 2. It also follows that the ratio of thermal to turbulent kinetic energy is

$$\frac{E_{\text{th}}}{\frac{1}{2} \rho \langle(\delta u_L)^2\rangle} \approx \frac{2\alpha}{3^{3/2}} \eta_{\text{turb}}^{-1} \approx \eta_{\text{turb}}^{-1} \quad (1)$$

Generation of magnetic field by small-scale dynamo processes in a turbulent flow follows from standard theory. In a high Reynolds number ( $\text{Re} > 10^3$ ) flow such as the ICM, an initial seed of vanishing strength<sup>22–27</sup> is amplified exponentially at the rate  $\Gamma = \sqrt{\text{Re}}/30\tau_L$ . After a short while ( $\propto \text{Re}^{-1/2}$ ), the magnetic field stops growing below the characteristic Alfvén scale,  $L_A \equiv v_A^3/C^{3/2}\epsilon_{\text{turb}}$ , where  $v_A = B/\sqrt{4\pi\rho}$  is the Alfvén speed and  $B$  the magnetic field strength, due to the back-reaction of magnetic tension. Nevertheless, magnetic energy

<sup>1</sup>Physics Department, ETH Zurich, Wolfgang-Pauli-Strasse 27, CH-8093 Zurich, Switzerland. <sup>2</sup>Nordita, KTH Royal Institute of Technology and Stockholm University, SE-10691 Stockholm, Sweden.



**Figure 1 | High-resolution simulation of a galaxy cluster in fully cosmological context.** **a, b**, Baryonic gas in the large-scale structure of the universe (**a**; bright is high, dark is low), and a zoom-in on the GC centre where numerical resolution is highest (**b**; colour-map inverted). Dynamic range of density (in  $\text{cm}^{-3}$ ) is  $\sim 10^6$ . The black dashed line marks the virial radius,  $R_{\text{vir}}$ .

continues to grow at the expense of turbulent kinetic energy as  $L_A$ , which marks the equipartition scale between kinetic and magnetic energy, shifts towards larger values<sup>5</sup>. Growth, however, is now proportional to the turbulent dissipation rate instead of being exponential with time. It is in this latter stage that the dynamo spends most of the time<sup>8</sup>. Recent state-of-the-art numerical work finds that for statistically isotropic and homogeneous turbulence, as found for the solenoidal component of the ICM<sup>9,20</sup>, the efficiency of the conversion of turbulent (kinetic) to magnetic energy is a universal number,  $C_E = 4\text{--}5\%$  (ref. 8).

Therefore, the evolution of magnetic energy in the ICM can be expressed in terms of the turbulence dissipation history as  $E_B(t) \equiv B^2/8\pi = C_E \int d\tau \rho \epsilon(\tau)$ . Combined with the above finding about  $\eta_{\text{turb}}$ , this leads to simple but significant expressions relating the fundamental properties of magnetic field and turbulence in the ICM. In fact, since turbulence dissipation contributes a constant fraction,  $\eta_{\text{turb}} \approx 1/3$ , of the ICM thermal energy, the ratio  $\beta_{\text{plasma}}$  of thermal pressure,  $P_{\text{gas}}$ , to magnetic energy can be written as

$$\beta_{\text{plasma}} \equiv \frac{P_{\text{gas}}}{B^2/8\pi} = \frac{\eta_{\text{turb}}^{-1}(\gamma-1)}{C_E} = 40 \left( \frac{\eta_{\text{turb}}}{1/3} \right)^{-1} \left( \frac{C_E}{0.05} \right)^{-1} \quad (2)$$

This means that for a massive GC,  $\beta_{\text{plasma}}$  is a constant, which depends neither on the specifics of the ICM conditions including turbulence, nor on the GC mass or age. It is instead simply determined by two fundamental parameters,  $C_E$  and  $\eta_{\text{turb}}$ , which describe the efficiency of magnetic field generation by the turbulent dynamo and of turbulent heating in structure formation, respectively. This is shown in Fig. 3a, where  $\beta_{\text{plasma}}$  is plotted as a function of cosmic time and exhibits 25% r.m.s. fluctuations, which should also characterize massive cluster-to-cluster variations. We can also compute the Alfvén scale. Since the turbulence is non-stationary and the magnetic properties respond on a timescale of the order of 1–2 eddy turnover times<sup>20</sup> we average the

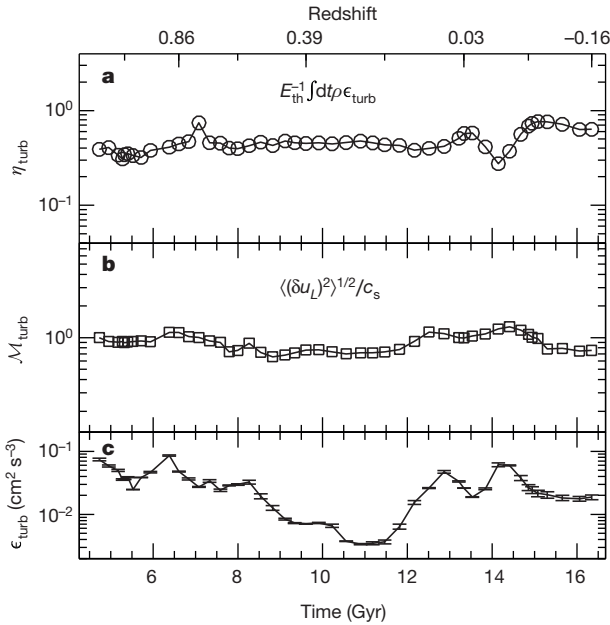
enclosing the volume that has nominally reached dynamical equilibrium. **c**, Vorticity magnitude on a scale twice the finest mesh size ( $\sim 20$  kpc). **d**, Temperature map. Complexity is due to shocks and contact discontinuities in the turbulent flow. Dynamic range of temperature (in K) and vorticity (in  $H_0^{-1}$ ) is  $\sim 10^3$ .

dissipation rate  $\epsilon_{\text{turb}}$  over 2 Gyr when calculating  $L_A$ . Expressing  $L_A$  in units of the turbulence outer scale we write

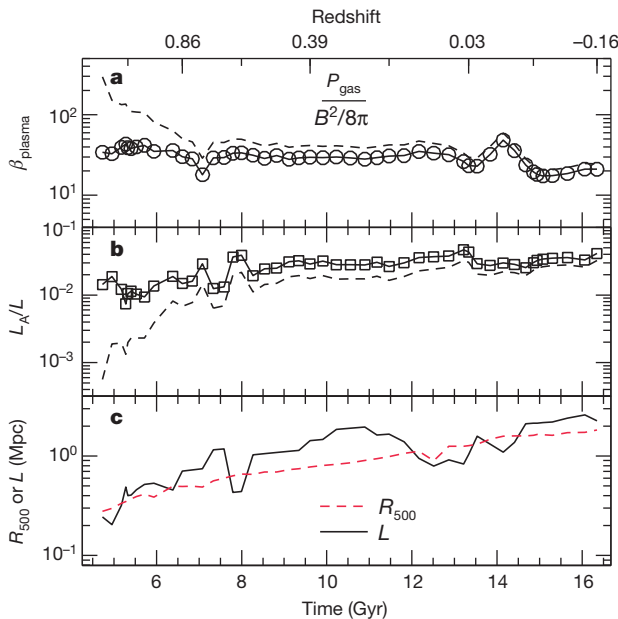
$$\begin{aligned} \frac{L_A}{L} &\equiv \frac{v_A^3}{C^{3/2} \langle \epsilon_{\text{turb}} \rangle L} = \left( \frac{3}{\gamma \beta_{\text{plasma}}} \right)^{3/2} \frac{c_s^3}{\langle (\delta u_L)^2 \rangle^{3/2}} \\ &= \frac{1}{100} \left( \frac{\beta_{\text{plasma}}}{40} \right)^{-3/2} \left( \frac{\mathcal{M}_{\text{turb}}}{1} \right)^{-3} \end{aligned} \quad (3)$$

We have already shown that both  $\beta_{\text{plasma}}$  and the turbulence Mach number remain constant during the evolution of a massive and dynamically young GC. Therefore, for such GCs, the Alfvén scale too remains a constant fraction of the turbulence driving scale, independent of time, GC mass and ICM conditions. In addition, given the large value of  $\beta_{\text{plasma}}$ , and that  $\mathcal{M}_{\text{turb}} \approx 1$ ,  $L_A$  is small compared to  $L$ . The temporal evolution of  $L_A/L$  is shown in Fig. 3b (see Extended Data Fig. 2 for a typical ICM spectrum of hydromagnetic turbulence). Finally, Fig. 3c shows that the evolution of the turbulent injection scale,  $L$ , closely follows that of  $L_A$  while tracking the growing characteristic scale of the GC ( $R_{500} = 0.5R_{\text{vir}}$ ) also plotted in the same panel. Note that the modulation of  $L_A$  reflects the changing turbulent conditions in the ICM and, in particular, is anti-correlated with  $\epsilon_{\text{turb}}$ , as generally expected.

The large value of  $\beta_{\text{plasma}}$  indicates that magnetic energy, like turbulent energy, is small compared to thermal energy ( $E_{\text{th}} \gg E_B$ ). Moreover, the small value of  $L_A/L$  indicates that the dynamo is far from saturation, and magnetic energy is also small in comparison to the turbulent kinetic energy ( $E_{\text{turb}} \gg E_B$ ). This energy hierarchy is fundamentally due to the efficiency  $\eta_{\text{turb}}$  with which turbulent energy is generated during gravitational collapse and the fraction  $C_E$  thereof that is converted into magnetic energy, namely  $E_{\text{th}}:E_{\text{turb}}:E_B = 1:\eta_{\text{turb}}:C_E\eta_{\text{turb}}$ . The values of  $\beta_{\text{plasma}}$  and  $L_A$  are in good agreement with recent measurements of magnetic field properties in GCs<sup>14–18</sup>. Here, they emerge from pure numerical modelling of



**Figure 2 | Temporal evolution of turbulence.** **a**,  $\eta_{\text{turb}}$ , the ratio of ICM thermal energy contributed by turbulent dissipation,  $\int_0^t dt \rho \epsilon_{\text{turb}}$ , to the total thermal energy,  $E_{\text{th}}$ . **b**, Turbulent Mach number, the ratio of the turbulent r.m.s. velocity  $\langle (\delta u_L)^2 \rangle^{1/2}$  to the sound speed  $c_s$ . **c**, Volumetric turbulent dissipation rate,  $\epsilon_{\text{turb}}$  (ref. 20). Error bars, variance of  $\epsilon_{\text{turb}}$ . All quantities are computed within 1/3 of the virial radius. Time in Gyr is reported on the bottom x axis, and the cosmological redshift on the top x axis.



**Figure 3 | Temporal evolution of magnetic field.** **a**,  $\beta_{\text{plasma}} = P_{\text{gas}} / (B^2 / 8\pi)$ , the ratio of ICM thermal to magnetic pressure, where  $B^2 / 8\pi = C_E \int dt \rho \epsilon_{\text{turb}}(\tau)$ . **b**,  $L_A / L$ , the ratio of Alfvén scale to the turbulent injection scale.  $L_A(\tau) = v_A^3 / [C^{3/2} \langle \epsilon_{\text{turb}} \rangle]$ , where  $v_A = B / \sqrt{4\pi}$  is the Alfvén speed and  $\langle \epsilon_{\text{turb}} \rangle$  is the turbulent dissipation rate smoothed over  $\tau = 2$  Gyr with a Gaussian filter. Quantities in **a** and **b** refer to a volume inside 1/3 of the virial radius. The dashed lines show transients to the asymptotic regime for an artificial  $t_{\text{start}} = 4.5$  Gyr. **c**, Turbulence injection scale  $L$  (solid black line) and the characteristic cluster size  $R_{500} = 0.5 R_{\text{vir}}$  (dashed line), enclosing a mass over-density of 500. Time in Gyr (bottom x axis) and the cosmological redshift (top x axis) are reported.

structure formation turbulence and MHD dynamo action, in the sense that they are found to derive their values from the parameters  $\eta_{\text{turb}}$  and  $C_E$ , which are determined numerically and not through parametric fits. Furthermore, the above energy hierarchy appears to remain unchanged during the evolution of the GC, and the turbulent dynamo in the ICM is as far from saturation today as it has virtually always been in the past. Figure 3c shows that the GC size and, therefore, its mass constantly grow. This implies that the gravitational potential energy and, therefore, the ICM thermal energy and turbulent energy also continue to grow. Meanwhile dynamo action tries to bring magnetic and turbulent energy into equipartition. Since all of these forms of energy grow simultaneously at the expense of the same gravitational energy, but with different constant efficiencies, their ratio remains unchanged, reflecting the value of those intrinsic efficiencies.

In other words, both  $\beta_{\text{plasma}}$  and  $L_A / L$  encode the efficiency of turbulence generation in structure formation and the efficiency of dynamo action. Therefore, they allow us to relate magnetic field observations in massive GCs to such properties of structure formation. This is in sharp contrast with other astrophysical bodies<sup>28–30</sup>—for example, the interstellar medium of galaxies, stars and compact objects—where the turbulence dynamo has long saturated and such information has been lost forever.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 22 January; accepted 27 April 2015.

1. Miniati, F. *et al.* Properties of cosmic shock waves in large-scale structure formation. *Astrophys. J.* **542**, 608–621 (2000).
2. Vazza, F., Brunetti, G., Gheller, C., Brunino, R. & Brüggén, M. Massive and refined. II. The statistical properties of turbulent motions in massive galaxy clusters with high spatial resolution. *Astron. Astrophys.* **529**, A17 (2011).
3. Miniati, F. The Matryoshka run: a Eulerian refinement strategy to study the statistics of turbulence in virialized cosmic structures. *Astrophys. J.* **782**, 21 (2014).
4. Zhuravleva, I. *et al.* Turbulent heating in galaxy clusters brightest in X-rays. *Nature* **515**, 85–87 (2014).
5. Schlüter, A. & Biermann, I. Interstellare Magnetfelder. *Z. Naturforsch. A* **5**, 237 (1950).
6. Cho, J., Vishniac, E. T., Beresnyak, A., Lazarian, A. & Ryu, D. Growth of magnetic fields induced by turbulent motions. *Astrophys. J.* **693**, 1449–1461 (2009).
7. Ryu, D., Kang, H., Cho, J. & Das, S. Turbulence and magnetic fields in the large-scale structure of the Universe. *Science* **320**, 909–912 (2008).
8. Beresnyak, A. Universal nonlinear small-scale dynamo. *Phys. Rev. Lett.* **108**, 035002 (2012).
9. Miniati, F. The Matryoshka run. II. Time-dependent turbulence statistics, stochastic particle acceleration, and microphysics impact in a massive galaxy cluster. *Astrophys. J.* **800**, 60 (2015).
10. Harrison, E. R. Fluctuations at the threshold of classical cosmology. *Phys. Rev. D* **1**, 2726–2730 (1970).
11. Kaiser, N. Evolution and clustering of rich clusters. *Mon. Not. R. Astron. Soc.* **222**, 323–345 (1986).
12. Navarro, J. F., Frenk, C. S. & White, S. D. M. A universal density profile from hierarchical clustering. *Astrophys. J.* **490**, 493–508 (1997).
13. Moore, B. *et al.* Dark matter substructure within galactic halos. *Astrophys. J.* **524**, L19–L22 (1999).
14. Clarke, T. E., Kronberg, P. P. & Böhringer, H. A new radio-X-ray probe of galaxy cluster magnetic fields. *Astrophys. J.* **547**, L111–L114 (2001).
15. Guidetti, D. *et al.* The intracluster magnetic field power spectrum in Abell 2382. *Astron. Astrophys.* **483**, 699–713 (2008).
16. Bonafede, A. *et al.* The Coma cluster magnetic field from Faraday rotation measures. *Astron. Astrophys.* **513**, A30 (2010).
17. Govoni, F. *et al.* Rotation measures of radio sources in hot galaxy clusters. *Astron. Astrophys.* **522**, A105 (2010).
18. Kuchar, P. & Enßlin, T. A. Magnetic power spectra from Faraday rotation maps. REALMAF and its use on Hydra A. *Astron. Astrophys.* **529**, A13 (2011).
19. Miniati, F. & Colella, P. Block structured adaptive mesh and time refinement for hybrid, hyperbolic + *N*-body systems. *J. Comput. Phys.* **227**, 400–430 (2007).
20. Beresnyak, A. & Miniati, F. Turbulent amplification and structure of intracluster magnetic field. *Astrophys. J.* (submitted).
21. Landau, L. D. & Lifshitz, E. M. *Fluid Mechanics* 2nd edn (Course of Theoretical Physics, Butterworth, 1987).
22. Subramanian, K., Narasimha, D. & Chitre, S. M. Thermal generation of cosmological seed magnetic fields in ionization fronts. *Mon. Not. R. Astron. Soc.* **271**, L15 (1994).
23. Kulsrud, R. M., Cen, R., Ostriker, J. P. & Ryu, D. The protogalactic origin for cosmic magnetic fields. *Astrophys. J.* **480**, 481–491 (1997).



24. Miniati, F. & Bell, A. R. Resistive magnetic field generation at cosmic dawn. *Astrophys. J.* **729**, 73 (2011).
25. Gregori, G. *et al.* Generation of scaled protogalactic seed magnetic fields in laser-produced shock waves. *Nature* **481**, 480–483 (2012).
26. Bertone, S., Vogt, C. & Enßlin, T. Magnetic field seeding by galactic winds. *Mon. Not. R. Astron. Soc.* **370**, 319–330 (2006).
27. Donnert, J., Dolag, K., Lesch, H. & Müller, E. Cluster magnetic fields from galactic outflows. *Mon. Not. R. Astron. Soc.* **392**, 1008–1021 (2009).
28. Stephens, I. W. *et al.* Spatially resolved magnetic field structure in the disk of a T Tauri star. *Nature* **514**, 597–599 (2014).
29. Bernet, M. L., Miniati, F., Lilly, S. J., Kronberg, P. P. & Dessauges-Zavadsky, M. Strong magnetic fields in normal galaxies at high redshift. *Nature* **454**, 302–304 (2008).
30. Zamaninasab, M., Clausen-Brown, E., Savolainen, T. & Tchekhovskoy, A. Dynamically important magnetic fields near accreting supermassive black holes. *Nature* **510**, 126–128 (2014).

**Acknowledgements** This work was supported by a grant from the Swiss National Supercomputing Center (CSCS) under project IDs S419 and S506.

**Author Contributions** F.M. carried out the cosmological simulations, computed the turbulence structure functions, derived equations (1), (2) and (3) and wrote most of the text. A.B. analysed the structure functions, testing the self-similar nature of second- and third-order structure functions within the inertial range and computing the dissipation rate. A.B. and F.M. computed the evolution of  $E_B$  and  $L_A$ .

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to F.M. (fm@phys.ethz.ch).

## METHODS

**Numerical model.** The simulation is carried out with CHARM, an adaptive-mesh-refinement cosmological code<sup>19</sup>. This code uses a directionally un-split variant of the piecewise parabolic method for hydrodynamics<sup>31</sup>, a constrained-transport algorithm for solenoidal MHD<sup>32</sup>, a time-centred modified symplectic scheme for the collisionless dark matter, and solves Poisson's equation with a second-order accurate discretization. The magnetic field remains negligible throughout, so the calculation is effectively hydrodynamic. For massive galaxy clusters, such as the Coma cluster, the ICM cooling time is a few times the age of the Universe<sup>33</sup>, so cooling and baryonic feedback processes are neglected. Heating of the intergalactic medium through photoionization is also neglected, with no consequences whatsoever for the generation of vorticity and turbulence at accretion shocks. We use a concordance  $\Lambda$ -CDM universe with normalized (in units of the critical value) total mass density,  $\Omega_m = 0.2792$ , baryonic mass density,  $\Omega_b = 0.0462$ , vacuum energy density,  $\Omega_\Lambda = 1 - \Omega_m = 0.7208$ , normalized Hubble constant  $h \equiv H_0/100 \text{ km s}^{-1} \text{ Mpc}^{-1} = 0.701$ , spectral index of primordial perturbation,  $n_s = 0.96$ , and r.m.s. linear density fluctuation within a sphere with a co-moving radius of  $8h^{-1} \text{ Mpc}$ ,  $\sigma_8 = 0.817$  (ref. 34). The simulated volume has co-moving size of  $L_{\text{Box}} = 240h^{-1} \text{ Mpc}$  on a side. The initial conditions are generated on three refinement levels with *grafic++* (made publicly available by D. Potter). For the coarsest level we use  $512^3$  co-moving cells, corresponding to a nominal spatial resolution of  $468.75h^{-1} \text{ co-moving kpc}$  and  $512^3$  particles of mass  $6.7 \times 10^9 h^{-1} M_\odot$  to represent the collisionless dark matter component. The additional levels allow for refined initial conditions in the volume where the galaxy cluster forms. The refinement ratio for both levels is  $n_{\text{ref}}^\ell \equiv \Delta x_\ell / \Delta x_{\ell+1} = 2$ ,  $\ell = 0, 1$ . Each refined level covers 1/8 of the volume of the next coarsest level with a uniform grid of  $512^3$  co-moving cells while the dark matter is represented with  $512^3$  particles. At the finest level the spatial resolution is  $\Delta x = 117.2h^{-1} \text{ co-moving kpc}$  and the particle mass is  $10^8 h^{-1} M_\odot$ . As the Lagrangian volume of the galaxy cluster shrinks under self-gravity, three additional uniform grids covering 1/8 of the volume of the next coarsest level are employed with  $512^3$ ,  $1,024^3$  and  $1,024^3$  co-moving cells, respectively, and  $n_{\text{ref}}^\ell = 2, 4, 2$ , for  $\ell = 2, 3, 4$ , respectively. All of them are in place by redshift 1.4, providing a spatial resolution of  $7.3h^{-1} \text{ co-moving kpc}$  in a region of  $7.5h^{-1} \text{ Mpc}$ , accommodating the whole virial volume of the GC. The ensuing dynamic range of resolved spatial scales is sufficiently large for the emergence of turbulence.

**GC characteristic quantities.** The GC and its formation history are reconstructed using our implementation of a HOP halo finder<sup>35</sup> and merger history code. The virial radius is defined as the region enclosing a mass over-density  $\Delta_c = 178\Omega_m^{0.45}$  with respect to the critical density<sup>36</sup>. At redshift  $z = 0$ , the virial radius is  $R_{\text{vir}} = 1.95h^{-1} \text{ Mpc}$ , and the corresponding enclosed mass,  $M_{\text{vir}} = 1.27 \times 10^{15} M_\odot$ . Also at  $z = 0$ , using  $\Delta_c = 500$  we find the characteristic radius  $R_{500} \approx 1h^{-1} \text{ Mpc}$ .

**Turbulence characterization.** The characteristic quantities describing the turbulence are inferred from the analysis of the structure functions. This analysis is

described in detail elsewhere<sup>3,9</sup>. Basically, we decompose the velocity into a solenoidal and a compressional component using a Hodge–Helmoltz decomposition, that is

$$\mathbf{v} = \mathbf{v}_s + \mathbf{v}_c, \quad \mathbf{v}_c = -\nabla\phi, \quad \mathbf{v}_s = \nabla \times \mathbf{A}, \quad \phi = \frac{1}{4\pi} \int \frac{\nabla \cdot \mathbf{v}}{r} d\mathbf{x}, \quad \mathbf{A} = \frac{1}{4\pi} \int \frac{\nabla \times \mathbf{v}}{r} d\mathbf{x} \quad (4)$$

and then we compute the second and third order structure functions of velocity increments of the solenoidal component,  $\delta v_i \equiv [\mathbf{v}_s(\mathbf{x} + \mathbf{l}) - \mathbf{v}_s(\mathbf{x})]_i$  (ref. 21)

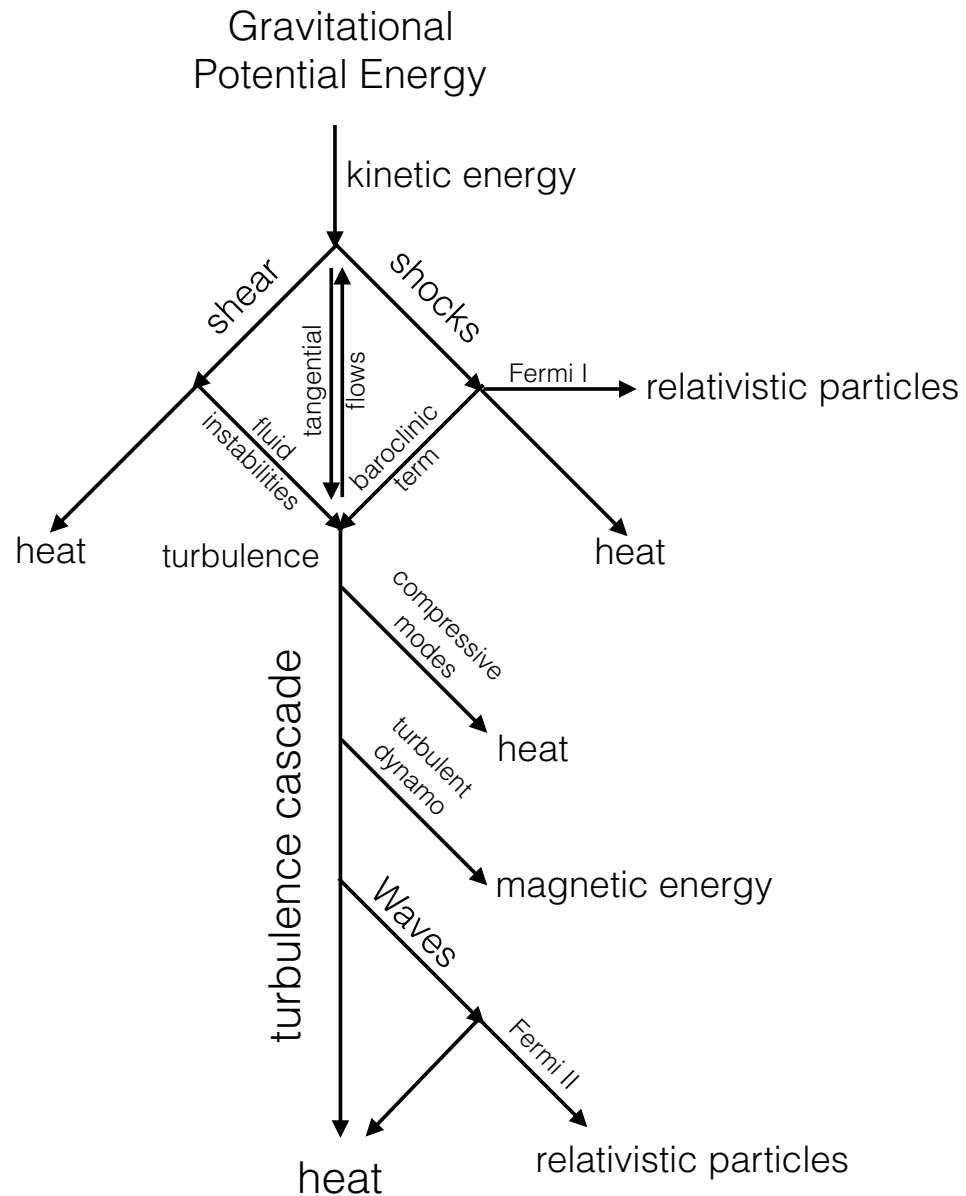
$$S_i(\mathbf{l}) \equiv \langle (\delta v_i)^p \rangle \quad (5)$$

where  $p = 2, 3$  indicates the structure function order, and  $i$  indicates the projection along or perpendicular to  $\mathbf{l}$ , for the longitudinal and transverse structure functions, respectively. To compute the structure functions we define sampling points randomly distributed inside the volume of interest (within 1/3 of the virial radius), and compute the velocity difference with respect to other randomly selected field points at a maximum distance of two virial radii. Once the velocity structure functions are computed, we define the velocity dispersion as half the asymptotic values of the second order structure function, and the outer scale as the separation at which that asymptotic value is reached. To compute the Mach number we divide the turbulent velocity dispersions by the sound speed,  $c_s = \sqrt{\gamma P/\rho}$ , computed by evaluating the mean value of each thermodynamic quantity within the same volume in which the sampling points are collected. Finally, the turbulent dissipation rate is computed by identifying the inertial range of the second and third order structure functions of the solenoidal velocity increments (for details see ref. 20).

**Code availability.** We have opted not to make the code available for practical reasons. However, the methods we adopt are published in the literature and are commonly used in the community. Amongst others, the publications mentioned in the above Methods section contain tests of our code against problems with known solutions and also with respect to solutions obtained with similar codes from independent authors.

31. Colella, P. Multidimensional upwind methods for hyperbolic conservation laws. *J. Comput. Phys.* **87**, 171–200 (1990).
32. Miniati, F. & Martin, D. F. Constrained-transport magnetohydrodynamics with adaptive mesh refinement in CHARM. *Astrophys. J.* **195** (Suppl.), 5 (2011).
33. Gaspari, M. & Churazov, E. Constraining turbulence and conduction in the hot ICM through density perturbations. *Astron. Astrophys.* **559**, A78 (2013).
34. Komatsu, E. et al. Five-year Wilkinson Microwave Anisotropy Probe observations: cosmological interpretation. *Astrophys. J.* **180** (Suppl.), 330–376 (2009).
35. Eisenstein, D. J. & Hut, P. HOP: a new group-finding algorithm for  $N$ -body simulations. *Astrophys. J.* **498**, 137–142 (1998).
36. Eke, V. R., Navarro, J. F. & Steinmetz, M. The power spectrum dependence of dark matter halo concentrations. *Astrophys. J.* **554**, 114–125 (2001).

# Diagram of ICM Turbulent Cascade

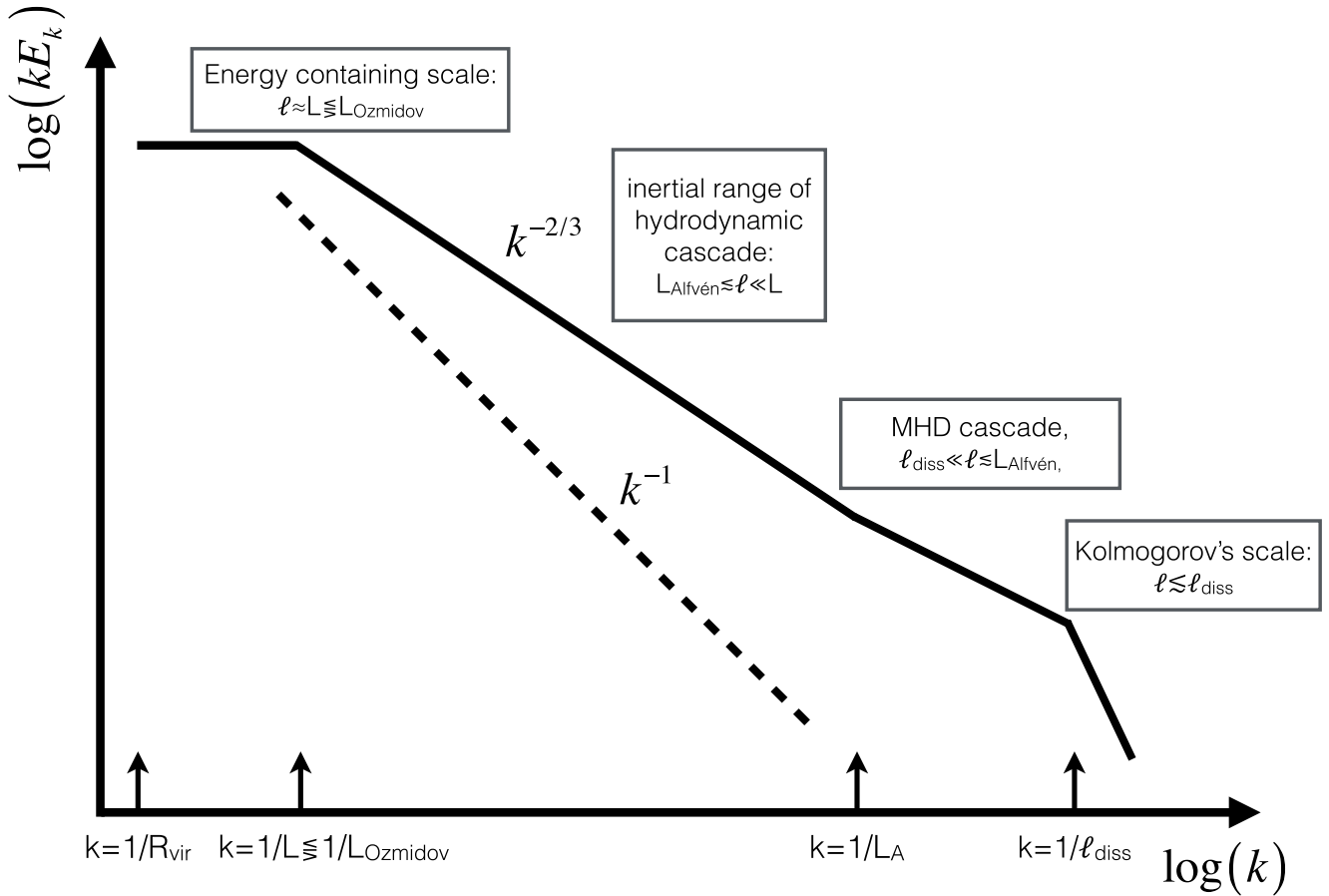


**Extended Data Figure 1 | Generation and cascade of ICM hydromagnetic turbulence.** First the gravitational potential energy is converted into kinetic energy of accretion flows. These generate shear and shocks which, in addition to dissipation, produce fluid instabilities and a baroclinic term, respectively, leading to turbulent flows. Shocks also accelerate particles via the Fermi I mechanism. Shocks do not dissipate tangential flows, which will either generate

turbulence, shear or shocks, or a combination thereof. The turbulence cascade includes dissipation of compressible modes at weak shocks, conversion of turbulent to magnetic energy via dynamo action, excitation of plasma waves accelerating relativistic particles via Fermi II mechanism, and viscous dissipation.



## Spectrum of ICM Hydromagnetic Turbulent Cascade



**Extended Data Figure 2 | Spectrum of ICM hydromagnetic turbulent cascade.** Characteristic spectrum of turbulent kinetic energy in the ICM. Solid and dashed lines correspond to the solenoidal (Kolmogorov-like) and the compressional (Burgers-like) velocity field, respectively. On the x axis, from left to right, we have marked the virial scale,  $R_{\text{vir}}$ , the injection scale,  $L$ , the

Ozmidov's scale,  $L_O$ , the Alfvén scale,  $L_A$ , and Kolmogorov's dissipation scale,  $\ell_{\text{diss}}$ . All quantities are time dependent and Ozmidov's scale is comparable to the injection scale, so at times turbulence in the radial direction could be suppressed by stratification.

# Large heterogeneities in comet 67P as revealed by active pits from sinkhole collapse

Jean-Baptiste Vincent<sup>1</sup>, Dennis Bodewits<sup>2</sup>, Sébastien Besse<sup>3</sup>, Holger Sierks<sup>1</sup>, Cesare Barbieri<sup>4</sup>, Philippe Lamy<sup>5</sup>, Rafael Rodrigo<sup>6,7</sup>, Detlef Koschny<sup>3</sup>, Hans Rickman<sup>8,9</sup>, Horst Uwe Keller<sup>10</sup>, Jessica Agarwal<sup>1</sup>, Michael F. A'Hearn<sup>2,11</sup>, Anne-Thérèse Auger<sup>5</sup>, M. Antonella Barucci<sup>12</sup>, Jean-Loup Bertaux<sup>13</sup>, Ivano Bertini<sup>14</sup>, Claire Capanna<sup>5</sup>, Gabriele Cremonese<sup>15</sup>, Vania Da Deppo<sup>16</sup>, Björn Davidsson<sup>6</sup>, Stefano Debei<sup>17</sup>, Mariolino De Cecco<sup>18</sup>, Mohamed Ramy El-Maarry<sup>19</sup>, Francesca Ferri<sup>14</sup>, Sonia Fornasier<sup>12</sup>, Marco Fulle<sup>20</sup>, Robert Gaskell<sup>21</sup>, Lorenza Giacomini<sup>14</sup>, Olivier Groussin<sup>5</sup>, Aurélie Guilbert-Lepoutre<sup>2</sup>, P. Gutierrez-Marques<sup>1</sup>, Pedro J. Gutiérrez<sup>22</sup>, Carsten Güttler<sup>1</sup>, Nick Hoekzema<sup>1</sup>, Sebastian Höfner<sup>1</sup>, Stubbe F. Hviid<sup>23</sup>, Wing-Huen Ip<sup>24</sup>, Laurent Jorda<sup>5</sup>, Jörg Knollenberg<sup>23</sup>, Gabor Kovacs<sup>1</sup>, Rainer Kramm<sup>1</sup>, Ekkehard Kührt<sup>23</sup>, Michael Küppers<sup>25</sup>, Fiorangela La Forgia<sup>4</sup>, Luisa M. Lara<sup>22</sup>, Monica Lazzarin<sup>4</sup>, Vicky Lee<sup>24</sup>, Cédric Leyrat<sup>12</sup>, Zhong-Yi Lin<sup>24</sup>, José J. Lopez Moreno<sup>22</sup>, Stephen Lowry<sup>26</sup>, Sara Magrin<sup>27</sup>, Lucie Maquet<sup>25</sup>, Simone Marchi<sup>28</sup>, Francesco Marzari<sup>27</sup>, Matteo Massironi<sup>29</sup>, Harald Michalik<sup>30</sup>, Richard Moissl<sup>25</sup>, Stefano Mottola<sup>23</sup>, Giampiero Naletto<sup>14,16,31</sup>, Nilda Oklay<sup>1</sup>, Maurizio Pajola<sup>14</sup>, Frank Preusker<sup>23</sup>, Frank Scholten<sup>23</sup>, Nicolas Thomas<sup>19</sup>, Imre Toth<sup>32</sup> & Cecilia Tubiana<sup>1</sup>

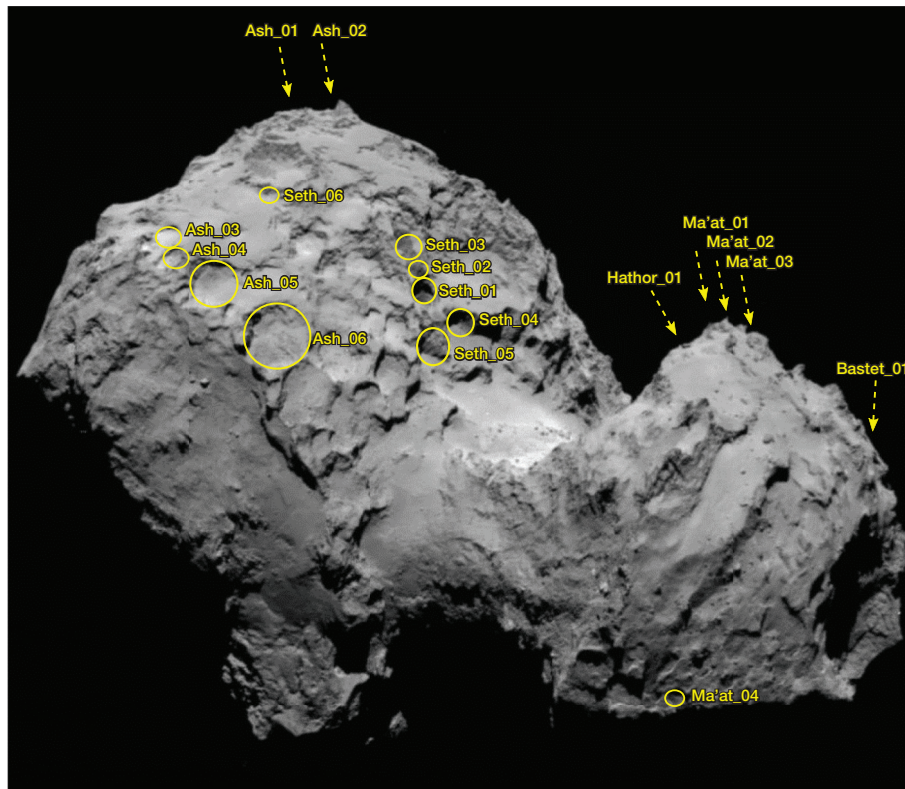
Pits have been observed on many cometary nuclei mapped by spacecraft<sup>1–4</sup>. It has been argued that cometary pits are a signature of endogenic activity, rather than impact craters such as those on planetary and asteroid surfaces. Impact experiments<sup>5,6</sup> and models<sup>7,8</sup> cannot reproduce the shapes of most of the observed cometary pits, and the predicted collision rates imply that few of the pits are related to impacts<sup>8,9</sup>. Alternative mechanisms like explosive activity<sup>10</sup> have been suggested, but the driving process remains unknown. Here we report that pits on comet 67P/Churyumov–Gerasimenko are active, and probably created by a sinkhole process, possibly accompanied by outbursts. We argue that after formation, pits expand slowly in diameter, owing to sublimation-driven retreat of the walls. Therefore, pits characterize how eroded the surface is: a fresh cometary surface will have a ragged structure with many pits, while an evolved surface will look smoother. The size and spatial distribution of pits imply that large heterogeneities exist in the physical, structural or compositional properties of the first few hundred metres below the current nucleus surface.

Understanding the differences in local activity of comet nuclei helps us to constrain how their surfaces have evolved since their formation. From July to December 2014, the OSIRIS (Optical, Spectroscopic, and Infrared Remote Imaging System) cameras on board Rosetta<sup>11</sup> continuously monitored the activity of comet 67P/Churyumov–Gerasimenko (referred to, hereafter, as comet 67P) from about a 30 km distance from the surface of the nucleus and resolved the fine structure of dust jets<sup>12</sup>. By means of stereo reconstruction, we found that broad jets can be separated into narrower features, which are linked unambiguously to

quasi-circular depressions and to walls of alcoves that are a few tens to a few hundreds of metres in diameter. These pits are remarkably symmetric and similar in size, and show interesting morphological details such as horizontal layers and terraces, vertical striations, and a smooth floor seemingly covered with dust. Some of these pits are as deep as a few hundred metres and provide a glimpse well below the nucleus surface. We detected a set of 18 quasi-circular pits on the northern hemisphere of comet 67P (Extended Data Table 1, Fig. 1). We observed that pits tend to cluster in small groups, and that several pits are active (Fig. 2). We measured the depth-to-diameter ratio ( $d/D$ ) of the pits and found that active pits have a high  $d/D = 0.73 \pm 0.08$ , while pits that are currently inactive are much shallower with mean  $d/D = 0.26 \pm 0.08$  (Extended Data Table 1, Fig. 3). The  $d/D$  ratio of these active pits is much higher than that of circular depressions on other comets:  $d/D = 0.1$  on comet 9P/Tempel 1 (ref. 4), and  $d/D = 0.2$  on comet 81P/Wild 2 (refs 13 and 14).

The difference in pit morphology on the three comets may reflect their different histories. For Jupiter family comets, the time since the last encounter with Jupiter is a proxy for the thermal history of the surface. Comet 9P is considered to be more processed by sublimation than comet 81P (ref. 3). In that view, comet 67P is relatively unprocessed by sublimation because its perihelion was brought from 2.7 astronomical units (AU) to 1.2 AU by a close encounter with Jupiter in 1959 (see Methods subsection ‘Orbit integration’). Comet 81P is also considered a young comet, but its pitted terrains are exposed to the Sun at perihelion and so have experienced much stronger erosion than the pitted areas on comet 67P even though it has spent less time in the

<sup>1</sup>Max-Planck-Institut für Sonnensystemforschung, Justus-von-Liebig-Weg 3, 37077 Göttingen, Germany. <sup>2</sup>University of Maryland, Department of Astronomy, College Park, Maryland 20742-2421, USA. <sup>3</sup>Scientific Support Office, European Space Research and Technology Centre/ESA, Keplerlaan 1, Postbus 299, 2201 AZ Noordwijk ZH, The Netherlands. <sup>4</sup>University of Padova, Department of Physics and Astronomy, Vicolo dell'Osservatorio 3, 35122 Padova, Italy. <sup>5</sup>Laboratoire d'Astrophysique de Marseille, UMR 7326, CNRS and Aix Marseille Université, 13388 Marseille Cedex 13, France. <sup>6</sup>Centro de Astrobiología, CSIC-INTA, 28850 Torrejón de Ardoz, Madrid, Spain. <sup>7</sup>International Space Science Institute, Hallerstraße 6, 3012 Bern, Switzerland. <sup>8</sup>Department of Physics and Astronomy, Uppsala University, Box 516, 75120 Uppsala, Sweden. <sup>9</sup>PAS Space Research Center, Bartyccka 18A, 00716 Warszawa, Poland. <sup>10</sup>Institut für Geophysik und extraterrestrische Physik (IGEP), Technische Universität Braunschweig, Mendelssohnstraße 3, 38106 Braunschweig, Germany. <sup>11</sup>Akademie der Wissenschaften zu Göttingen and Max-Planck-Institut für Sonnensystemforschung, Justus-von-Liebig-Weg 3, 37077 Göttingen, Germany. <sup>12</sup>LESIA-Observatoire de Paris, CNRS, Université Pierre et Marie Curie, Université Paris Diderot, 5 place Jules Janssen, 92195 Meudon, France. <sup>13</sup>LATMOS, CNRS/UVSQ/IPSL, 11 boulevard d'Alembert, 78280 Guyancourt, France. <sup>14</sup>Centro di Ateneo di Studi ed Attività Spaziali “Giuseppe Colombo” (CISAS), University of Padova, via Venezia 15, 35131 Padova, Italy. <sup>15</sup>INAF, Osservatorio Astronomico di Padova, Vicolo dell'Osservatorio 5, 35122 Padova, Italy. <sup>16</sup>CNR-IFN UOS Padova LUXOR, via Trasea 7, 35131 Padova, Italy. <sup>17</sup>Department of Industrial Engineering, University of Padova, via Venezia 1, 35131 Padova, Italy. <sup>18</sup>University of Trento, via Mesiano 77, 38100 Trento, Italy. <sup>19</sup>Physikalisches Institut der Universität Bern, Sidlerstraße 5, 3012 Bern, Switzerland. <sup>20</sup>INAF Osservatorio Astronomico, via Tiepolo 11, 34014 Trieste, Italy. <sup>21</sup>Planetary Science Institute, Tucson, Arizona 85719, USA. <sup>22</sup>Instituto de Astrofísica de Andalucía (CSIC), Glorieta de la Astronomía s/n, 18008 Granada, Spain. <sup>23</sup>Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Planetenforschung, Rutherfordstraße 2, 12489 Berlin, Germany. <sup>24</sup>National Central University, Graduate Institute of Astronomy, 300 Chung-Da Rd, Chung-Li 32054, Taiwan. <sup>25</sup>Operations Department, European Space Astronomy Centre/ESA, PO Box 78, 28691 Villanueva de la Canada, Madrid, Spain. <sup>26</sup>The University of Kent, School of Physical Sciences, Canterbury, Kent CT2 7NZ, UK. <sup>27</sup>University of Padova, Department of Physics and Astronomy, via Marzolo 8, 35131 Padova, Italy. <sup>28</sup>Solar System Exploration Research Virtual Institute, Southwest Research Institute, 1050 Walnut Street, Suite 300, Boulder, Colorado 80302, USA. <sup>29</sup>Dipartimento di Geoscienze, University of Padova, via Giovanni Gradenigo 6, 35131 Padova, Italy. <sup>30</sup>Institut für Datentechnik und Kommunikationsnetze der Technische Universität Braunschweig, Hans-Sommer-Straße 66, 38106 Braunschweig, Germany. <sup>31</sup>University of Padova, Department of Information Engineering, via Gradenigo 6/B, 35131 Padova, Italy. <sup>32</sup>Konkoly Observatory of the Hungarian Academy of Sciences, PO Box 67, 1525 Budapest, Hungary.

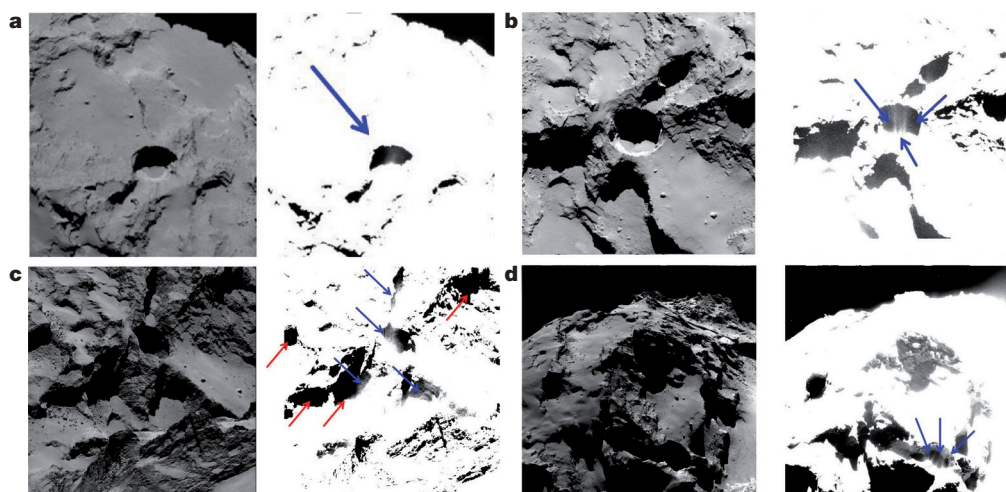


**Figure 1 | Location of the pits considered in this study.** A non-exhaustive catalogue of depressions sharing similar morphologies to those unambiguously linked to jets in the Seth and Ma'at regions.

inner Solar System. Deep active pits on comets are seemingly found preferentially on surfaces that have not been notably eroded.

The terrain morphology inside the pits on comet 67P is not uniform and is classified as: very smooth texture; fractured terrain, terraces and alcoves; or globular texture. The globular texture is detected only in the deep pits and at a few additional locations on the nucleus, where deeper near-surface layers can be observed. This morphology extends to a depth of at least 200 m below the current nucleus surface (see, for example, pit Seth\_01, Extended Data Fig. 2).

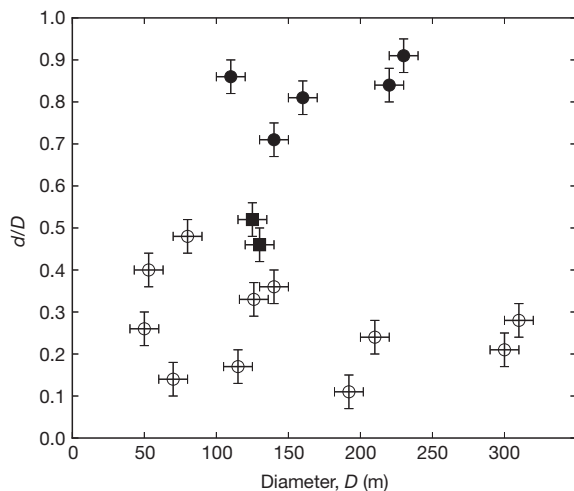
Jets arise from the edges of active pits, primarily from heavily fractured and globular morphologies (Fig. 2). However, the  $d/D$  ratio cannot be explained by current sublimation-driven retreat of the walls. Excavating a pit like Seth\_01 by sublimating ice on the wall and floor would take more than 7,000 years (Methods). The cylindrical shape of most pits also provides evidence against formation by erosion, because this would result in elongated shapes and a latitudinal dependence of the pit distribution on the surface.



**Figure 2 | Jet-like features in the Seth region.** a–d, Views of the main active pit in the Seth region, at different angles of solar illumination. The illuminated area of the pit is the south wall (a), the north wall (b), the east wall (c) and the southeast wall (d). Blue arrows point to detected jets; red arrows indicate areas

where no activity could be observed, either from the walls or from their surroundings. The left images are the original data; the right images are linearly stretched in brightness to display the lowest 5% of the intensity values.





**Figure 3 | Depth-to-diameter ratio as a function of pit diameter.** Filled symbols describe active pits; empty symbols describe currently inactive pits. Filled circles are active pits in the Seth region; filled squares are active pits in the Ma'at region. The lower value of  $d/D$  for pits in the latter might indicate a different formation mechanism. Error bars represent the uncertainties inherent to the shape reconstruction technique (stereo-photogrammetry) used to produce the digital terrain model of the comet<sup>12</sup>.

The 380 pits observed on comet 9P have been associated with explosive activity<sup>9</sup>. In the weeks before its encounter with this comet, the spacecraft Deep Impact observed at least 10 outbursts, the largest of which ejected an estimated  $(6\text{--}30) \times 10^4$  kg of material<sup>10,15,16</sup>. The observations suggest that these outbursts originated from a series of pits located in a belt around the nucleus. At 4.11 AU from the Sun, on 30 April 2014, OSIRIS observed an outburst on comet 67P (ref. 17). Depending on the assumed size distribution of the ejected dust, the resulting plume contained  $10^3\text{--}10^5$  kg of material, and was thus of similar magnitude to the outbursts observed on comet 9P. Such outbursts are too small to create the observed pits by explosive excavation. Assuming a constant density<sup>12</sup> of  $470 \text{ kg m}^{-3}$ , a typical large active pit on comet 67P would have contained approximately  $10^9$  kg of material,  $10^4$  times more than the upper limit on the mass of the material excavated by the observed outburst.

We propose that the pits are formed via sinkhole collapse, when the ceiling of a subsurface cavity becomes too thin to support itself (Fig. 4 and Methods). Because the size of sinkholes depends on the material strength of the top layers, sinkholes in a given terrain are all of similar size. They are characterized by circular depressions aligned with the local gravity vector<sup>18</sup>.

On cometary nuclei, the removal of subsurface volatiles may generate a void. Failure of the cavity's ceiling propagates upward. From the observed pit diameters and depths, and by treating the cavity's roof as

an unsupported beam failing under its own weight, we estimate that the collapsing layer has an average tensile strength of 50 Pa (Extended Data Fig. 8 and Methods). This value is similar to the lower-limit estimate based on overhangs on the surface<sup>19</sup>. The collapse exposes fresh material in the walls of the pit, which sublimates to produce the observed jets. Such collapse may very well be the driver of the 30 April 2014 outburst from comet 67P and the mini-outbursts from comet 9P. The morphology and expansion of the dust plume of the 30 April 2014 outburst from comet 67P suggest that most of the activity arose from an area within  $30^\circ$  of latitude of the north pole<sup>17</sup>, compatible with the location of the pits in the Seth region.

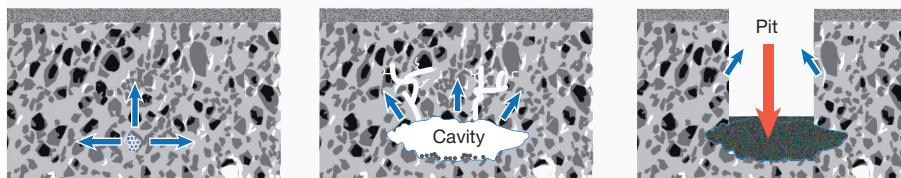
The collapse itself is a sudden event, but the cavity 100–200 m below the surface could have been growing over a much longer timescale. We explore three cavity formation scenarios: (1) primordial voids inherited from formation; (2) direct sublimation of super volatiles (CO and CO<sub>2</sub>) as an evolutionary process; and (3) deep subsurface sublimation triggered by a secondary source of energy.

(1) The primordial scenario implies that voids existed in the nucleus since its formation. This is possible if the comet formed by slow accretion of cometsimals of tens to hundreds of metres in size. Low collision speeds would prevent crushing the cavities<sup>20</sup>. A weakening of the surface due to direct sublimation would trigger roof collapse.

(2) Cavity formation can also be an evolutionary process. Because comet nuclei have very low thermal conductivity<sup>21</sup>, direct sublimation of hexagonal water ice at the required depths would occur at an extremely low rate and can therefore be ruled out. It is possible, however, to sublimate more volatile ices like CO and CO<sub>2</sub> at lower temperatures. The fact that we do not see pits everywhere suggests that these super volatiles may not be distributed evenly inside the nucleus; such heterogeneity has been observed on the surface of other comets (9P, ref. 22; 103P, ref. 23).

(3) A subsurface energy source may provide the heat necessary to sublimate a large cavity. A candidate is the phase transition in water ice from an amorphous to a crystalline structure. Crystallization has been used to explain many cometary activity features, and has been suggested as the underlying process for the distant outbursts of comet 1P/Halley and the chaotic behaviour of comet 29P/Schwassmann–Wachmann<sup>24</sup>, or the outburst of comet 17P/Holmes<sup>25</sup>. Different models have placed the crystallization front at depths ranging from a few metres to hundreds of metres<sup>26,27</sup>. We find that a subsurface cavity of the size of the observed pits would require the phase transition of at least 600 kg of amorphous ice, corresponding to a sphere of 2 m in diameter at most (see Methods). The detailed calculation of the amount needed is beyond the scope of this Letter.

Ultimately, regardless of the process creating the initial subsurface cavity, active pits indicate that large structural and/or compositional heterogeneities exist within the first few hundred metres below the current nucleus surface of comet 67P. Clusters of active pits and collapsed structures are signatures of former cavities underneath, and reflect the thermal history of the nucleus.



**Figure 4 | Pit formation mechanism by sinkhole collapse.** A typical comet surface with a layer of dust covering a mixture of dust and volatile material. A subsurface heat source sublimates surrounding ices. This gas then escapes or relocates, thus forming a cavity. When the ceiling becomes too thin to support

its own weight it collapses, creating a deep, circular pit with a smooth bottom. Newly exposed material in the pit's walls can start to sublimate. Blue arrows and white lines describe the escape of volatiles and fracturing of the surrounding material; red arrow shows the collapse of the cavity ceiling.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 15 February; accepted 6 May 2015.**

- Keller, H. U., Kramm, R. & Thomas, N. Surface features on the nucleus of comet Halley. *Nature* **331**, 227–231 (1988).
- Soderblom, L. A. *et al.* Observations of comet 19P/Borrelly by the miniature integrated camera and spectrometer aboard Deep Space 1. *Science* **296**, 1087–1091 (2002).
- Brownlee, D. E. *et al.* Surface of young Jupiter family comet 81P/Wild 2: view from the Stardust spacecraft. *Science* **304**, 1764–1769 (2004).
- Thomas, P. C. *et al.* The nucleus of comet 9P/Tempel 1: shape and geology from two flybys. *Icarus* **222**, 453–466 (2013).
- A'Hearn, M. F. *et al.* Deep Impact: excavating comet Tempel 1. *Science* **310**, 258–264 (2005).
- Schultz, P., Hermalyn, B. & Veverka, J. The Deep Impact crater on 9P/Tempel-1 from Stardust-NEXT. *Icarus* **222**, 502–515 (2013).
- Holsapple, K. A. & Housen, K. R. A crater and its ejecta: an interpretation of Deep Impact. *Icarus* **187**, 345–356 (2007).
- Vincent, J.-B., Oklay, N., Marchi, S., Höfner, S. & Sierks, H. Craters on comets. *Planet. Space Sci.* **107**, 53–63 (2015).
- Belton, M. J. S. *et al.* The origin of pits on 9P/Tempel 1 and the geologic signature of outbursts in Stardust-NEXT images. *Icarus* **222**, 477–486 (2013).
- Belton, M. J. S. *et al.* Cometary cryo-volcanism: source regions and a model for the UT 2005 June 14 and other mini-outbursts on comet 9P/Tempel 1. *Icarus* **198**, 189–207 (2008).
- Keller, H. U. *et al.* OSIRIS – The scientific camera system onboard Rosetta. *Space Sci. Rev.* **128**, 433–506 (2007).
- Sierks, H. *et al.* On the nucleus structure and activity of comet 67P/Churyumov-Gerasimenko. *Science* **347**, aaa1044 (2015).
- Kirk, R. *et al.* Topography of the 81P/Wild 2 nucleus derived from Stardust stereoisimages. *Lunar Planet. Sci.* **XXXVI**, 2244 (2005).
- Basilevsky, A. T. & Keller, H. U. Comet nuclei: morphology and implied processes of surface modification. *Planet. Space Sci.* **54**, 808–829 (2006).
- Farnham, T. L. *et al.* Dust coma morphology in the Deep Impact images of comet 9P/Tempel 1. *Icarus* **187**, 26–40 (2007).
- Feldman, P. D. *et al.* Hubble Space Telescope observations of comet 9P/Tempel 1 during the Deep Impact encounter. *Icarus* **187**, 113–122 (2007).
- Tubiana, C. *et al.* 67P/Churyumov-Gerasimenko: activity between March and June 2014 as observed from Rosetta/OSIRIS. *Astron. Astrophys.* **573**, A62 (2015).
- Waltham, T., Bell, F. G. & Culshaw, M. G. *Sinkholes and Subsidence* (Springer, 2007).
- Thomas, N. *et al.* The morphological diversity of comet 67P/Churyumov-Gerasimenko. *Science* **347**, aaa0440 (2015).
- Weissman, P. R., Asphaug, E. & Lowry, S. C. in *Comets II* (eds Festou, M. C. *et al.*) 337–357 (Univ. Arizona Press, 2004).
- Groussin, O. *et al.* The temperature, thermal inertia, roughness and color of the nuclei of comets 103P/Hartley 2 and 9P/Tempel 1. *Icarus* **222**, 580–594 (2013).
- Feaga, L. M., A'Hearn, M. F., Sunshine, J. M., Groussin, O. & Farnham, T. L. Asymmetries in the distribution of H<sub>2</sub>O and CO<sub>2</sub> in the inner coma of comet 9P/Tempel 1 as observed by Deep Impact. *Icarus* **190**, 345–356 (2007).
- A'Hearn, M. F. *et al.* EPOXI at comet Hartley 2. *Science* **332**, 1396–1400 (2011).
- Prialnik, D., Benkhoff, J. & Podolak, M. in *Comets II* (eds Festou, M. C. *et al.*) 359–387 (Univ. Arizona Press, 2004).
- Hillman, Y. & Prialnik, D. A quasi 3-D model of an outburst pattern that explains the behavior of comet 17P/Holmes. *Icarus* **221**, 147–159 (2012).
- Tancredi, G., Rickman, H. & Greenberg, J. M. Thermochemistry of cometary nuclei. I. The Jupiter family case. *Astron. Astrophys.* **286**, 659–682 (1994).
- Marboeuf, U. & Schmitt, B. How to link the relative abundances of gas species in coma of comets to their initial chemical composition? *Icarus* **242**, 225–248 (2014).

**Acknowledgements** OSIRIS was built by a consortium of the Max-Planck-Institut für Sonnensystemforschung, Katlenburg-Lindau, Germany, the CISAS, University of Padova, Italy, the Laboratoire d'Astrophysique de Marseille, France, the Instituto de Astrofísica de Andalucía, CSIC, Granada, Spain, the Research and Scientific Support Department of the European Space Agency, Noordwijk, The Netherlands, the Instituto Nacional de Técnica Aeroespacial, Madrid, Spain, the Universidad Politécnica de Madrid, Spain, the Department of Physics and Astronomy of Uppsala University, Sweden, and the Institut für Datentechnik und Kommunikationsnetze der Technischen Universität Braunschweig, Germany. The support of the national funding agencies of Germany (DLR), France (CNES), Italy (ASI), Spain (MEC), Sweden (SNSB), and the ESA Technical Directorate is acknowledged. This work was also supported by NASA JPL contract 1267923 to the University of Maryland (M.F.A'H. and D.B.). M.F.A'H. is a Gauss Professor at the Akademie der Wissenschaften zu Göttingen and Max-Planck-Institut für Sonnensystemforschung (Germany). This research has made use of NASA's Astrophysics Data System Bibliographic Services. We thank H. J. Melosh for reviews and criticism.

**Author Contributions** J.-B.V. led the study, identified the pits and measured their global parameters. D.B. analysed outbursts and phase change transitions and prepared the sinkhole model. S.B. performed the detailed morphology analysis. H.S., C.B., P.L., R.R., D.K. and H.R. are the lead scientists of the OSIRIS project. The other authors are co-investigators who built and ran this instrument and made the observations possible, and associates and assistants who participated in the study.

**Author Information** All data presented in this paper will be delivered to the ESA's Planetary Science Archive and NASA's Planetary Data System in accordance with the schedule established by the Rosetta project and will be available on request before that archiving. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.-B.V. ([vincent@mps.mpg.de](mailto:vincent@mps.mpg.de)).

## METHODS

**Detection of activity.** Cometary activity is typically defined as the ensemble of physical processes forming the gas and dust coma that escapes from the nucleus. The main driver of activity is the solar insolation, which triggers the sublimation of volatiles trapped in the subsurface of the nucleus<sup>28</sup>. The liberated gas expands into vacuum and drags along refractory grains from the surface. It has been known since the first *in situ* mission to a comet that this activity is not uniformly distributed over the nucleus although the reasons for this anisotropy are not well understood<sup>1</sup>.

From the uneven distribution of active sources on the surface, anisotropies in the coma arise in the form of narrow dusty streams (hereafter called 'jets'), which expand straight from the nucleus for at least some distance<sup>29</sup>. Neither their source nor the physics of their formation have been fully explained yet, although many authors have proposed some explanations such as patches of enhanced H<sub>2</sub>O ice content, localized super-volatile release from steep-sided pits, or repetitive mini-outbursts<sup>30</sup>. It is not clear whether these features are linked to volatiles at their footprint or if they trace the shock front between competing gas flows from nearby areas<sup>31</sup>.

In OSIRIS images, jets appear as fuzzy streams of bright material arising from specific areas on the nucleus surface. They are typically detected against the coma or a dark background, which can be either empty space or cast shadows. They are seen at all spatial scales, from large features spanning several tens of kilometres, down to the limit of spatial resolution. The smallest features detected so far are a few pixels across, which translates into a couple of metres at most. Their typical surface brightness is 10% to 40% higher than the surrounding background space, that is, the general coma<sup>12</sup>. By monitoring the activity and observing these jets from different angles we can perform stereo imaging, reconstruct their three-dimensional structure and trace them back precisely to morphological features on the surface.

**Orbit integration.** Observations and orbit reconstructions have shown that comet 67P had a close encounter with Jupiter that brought its perihelion from 2.7 AU to 1.2 AU, in 1959 (JPL Horizons ephemerides, <http://ssd.jpl.nasa.gov/horizons.cgi>). We reconstructed its orbit before that time, on the basis of a well-established integration model<sup>32</sup>. For the initial conditions and their errors, we refer to the database of IMCCE (<http://www.imcce.fr/langues/en/ephemerides/>). We compute 200 clone orbits with random Gaussian small variations of the initial conditions considering their Gaussian errors. From these 200 clone orbits, we deduce the mean perihelion distance and its standard deviation ( $\sigma$ ). We find that 84% of the orbits in the interval  $[(\text{mean} - \sigma), (\text{mean} + \sigma)]$  and the orbits beyond  $(\text{mean} + \sigma)$  have a perihelion distance greater than at least 2 AU with a mean value always greater than 3 AU (Extended Data Fig. 1).

**Morphology, variegation and activity of the pits.** The pit morphologies are presented in more detail in Extended Data Figs 3, 4. The complete list of OSIRIS images used for this study is given in Extended Data Table 2.

The activity identified in Seth\_01 covers the portion of the pit presented in Fig. 1, which displays different morphologies and textures. Therefore, it is not clear at this point that a specific texture and morphology is linked to the active pits. The detailed observations of the pits Ma'at\_01 and Ma'at\_02 seem to indicate that heavily fractured terrains are, however, favourably associated with activity. Extended Data Fig. 3 highlights the multiple joints that are also associated with the globular texture for Ma'at\_01. Thus, fractured texture might be favourable for these active pits, probably because it allows the heat to propagate deeper into the interior and sublimate the ices. One other possible location for the activity inside the pits could be the terraces seen in Seth\_01 (and maybe in Ma'at\_02, although they are less developed). The two terraces highlighted in Extended Data Fig. 3 cover around 50% of the circumference of the Seth\_01 pit, and they match the 50% where activity has been identified so far. Therefore, the terraces could be the source of the activity if they expose some kind of fresh ice (or gas/ice from the coma falling back and depositing on this flat surface).

The contact between the edges of the pits and the surroundings is different between the active pits Seth\_01 and Ma'at\_01. This could be the result of different mechanisms that formed them or the primordial morphology of the region. The bottoms of most pits are covered with a fine dusty material and boulders, which could be an indication of the relative age of these pits. Seth\_01's floor appears very flat (Extended Data Figs 3, 4), with a very smooth structure that does not contain any boulders. The floors of Seth\_02 and Seth\_03, pits where activity has been identified, share the same textural characteristics as Seth\_01. The viewing conditions are less favourable for Ma'at\_01; however, Extended Data Figs 3–5 show few boulders, all of small size. The same figures highlight that Ma'at\_02 has a much higher number of boulders with larger sizes. These boulders may be an indirect way of estimating the relative age of the pits, because boulders accumulate with time. Thus, boulder-free floors represent the youngest pits. The relative age dating of these pits could also be speculated from the Ma'at\_01 to Ma'at\_03 series of pits.

With Ma'at\_01 being the youngest and Ma'at\_03 the oldest, one can see the degradation of the wall of the pits and the accumulation of material within the pit. The accumulation of boulders is rather limited in Ma'at\_03, although the degradation of the rim is in a more advanced stage when compared to the other two, which confirms that it is the oldest. This low accumulation could be due to the geometry of Ma'at\_03 or related to the original depth of the pit, which is most likely to have been smaller. The boulder-size distributions in the Seth and Ma'at pits are shown in Extended Data Figs 5, 6.

We used additional images obtained through filters near the visible spectrum (blue: 480 nm, orange: 649 nm, infrared: 989 nm) to see if, in addition to the peculiar morphology, pits present a different colour to the rest of the surface. By using filter ratios to limit the effect of topography and illumination conditions, we found that the floor and walls of the pits exhibit the same less-red spectral slope as the active Hapi region (Extended Data Fig. 7). If we denote the reflectance by  $R$ , then we measure a ratio  $R_{\text{infrared}}/R_{\text{blue}} = 1.8$  in the active area (pits) and  $R_{\text{infrared}}/R_{\text{blue}} = 2.1$  elsewhere on the nucleus. A full understanding of the implications of the compositional differences within the nucleus will require a dedicated investigation, but the difference in spectral slope observed in Extended Data Fig. 7 already indicates that spectral variation is an intrinsic property of currently active regions on comet 67P.

**Pit growth.** A major question is whether the  $d/D$  ratio of the pits can be explained by the current sublimation-driven retreat of the walls. We see jets arising from the edges of active pits (Fig. 1), indicating that erosion currently does occur. We first consider slowly excavating a pit by sublimating subsurface ice on the walls and floor and growing the depressions in both diameter and depth. We take as an example the most active pit (Seth\_01). With a diameter of 220 m and a depth of 185 m, it has a volume of  $7 \times 10^9 \text{ m}^3$ , which corresponds to  $3.3 \times 10^9 \text{ kg}$  of material if we assume a constant density of  $470 \text{ kg m}^{-3}$ . Current models of activity for comet 67P (refs 33–35) describe a global dust production rate of  $9.3 \text{ kg s}^{-1}$  at 3.5 AU, which translates into only  $15 \text{ g s}^{-1}$  of dust emitted from a single pit. Additionally, the varying latitudes and seasons limit the pits' illumination to only a few hours per comet day for the walls. In some cases, the pit floor is only barely illuminated, if at all. Considering that most currently observed pits will be in polar night at perihelion and will not experience many changes in dust production rate, it would take more than 7,000 years to dig out one pit.

Erosion is a second-order process that will slowly modify the pits after they are formed. This is supported by our observations; several active pits display alcoves within their walls, which we interpret as signatures of continued growth as a result of erosion by sublimation, block falls and wall retreats long after the pit formation, because these alcoves are always facing the direction of most insolation received per comet rotation.

**Phase transition.** Crystallization has been invoked to explain many cometary activity features, and has been suggested as the underlying process for the distant outbursts of comet 1P/Halley<sup>24</sup>, the chaotic behaviour of comet 29P/Schwassmann–Wachmann<sup>24</sup> and the violent outburst of comet 17P/Holmes<sup>25</sup>.

From the ratio between the latent heat of the amorphous-to-crystalline transition ( $9 \times 10^4 \text{ J kg}^{-1}$ , refs 36, 37) and of the sublimation of hexagonal ice ( $0.334 \text{ J kg}^{-1}$ ), the phase transition of 1 kg of amorphous ice to crystalline ice provides enough energy to sublimate 270 kg of hexagonal ice, provided that crystallization occurs on a timescale short enough for the phase transition to effectively heat the surrounding ice. Using typical low thermal inertia, Marboeuf and Schmitt<sup>28</sup> find that crystallization proceeds to a depth of only approximately 1 m. Other studies estimate that the crystallization front should extend to depths of between about 5 m and about 80 m (ref. 37), or much greater depths<sup>37</sup>.

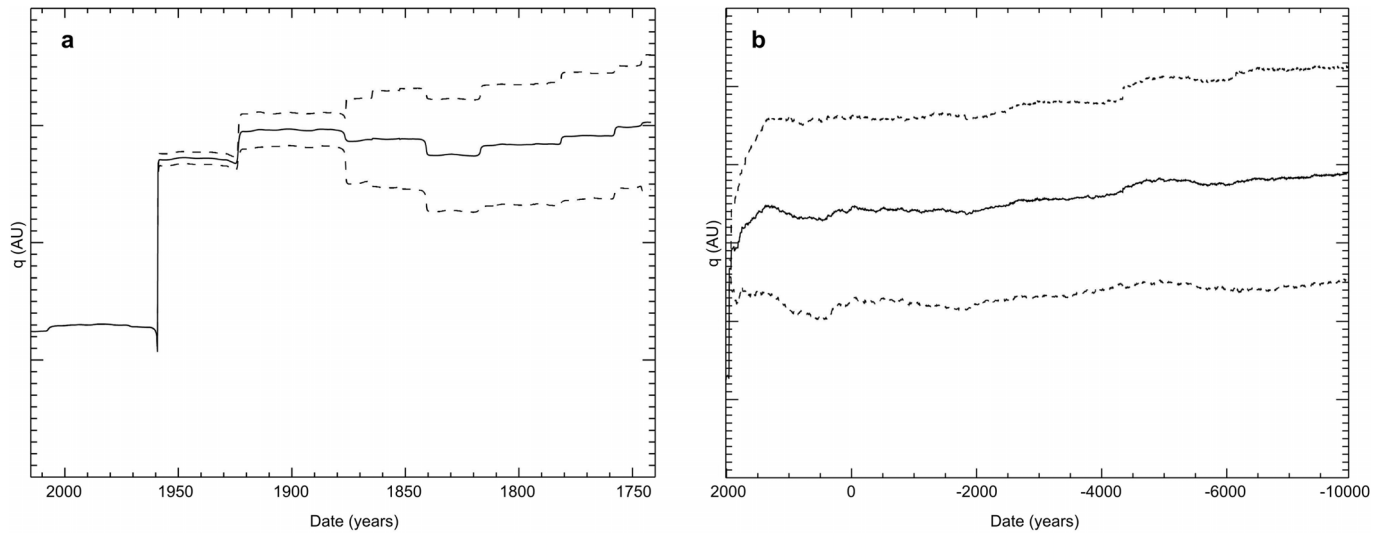
Given the chaotic orbital evolution of comet 67P, we estimate that a 100 m deep layer could have recently reached the appropriate characteristics (100–120 K leading to a phase transition on a timescale of months to a year) only if the local thermal inertia is high ( $250 \text{ J m}^{-2} \text{ K}^{-1} \text{ s}^{-1/2}$  and above), a value more than five times what has been measured on comet 67P. For lower values of the thermal inertia, the phase transition can occur at a depth of 100–200 m only after a long period of time in the inner Solar System. A cavity could have formed much earlier in the history of the comet, even if the final collapse that produced the observed sinkhole occurred only recently. A subsurface cavity the size of the pits we observe would require the phase transition of approximately 600 kg of amorphous ice to crystalline ice. If we assume a density of  $470 \text{ kg m}^{-3}$  and a porosity of 70–80%, we obtain a 20–40% ice mass fraction in the nucleus (ice density is  $920 \text{ kg m}^{-3}$ , solid material is half silicate and half organics, with respective densities of  $3,500 \text{ kg m}^{-3}$  and  $2,200 \text{ kg m}^{-3}$ ). Therefore, 600 kg of ice would be embedded in 1,500–3,000 kg of cometary material and would occupy a volume of 3–6 m<sup>3</sup>, that is, a sphere of at most 2 m in diameter. Upon experiencing its phase transition, this pocket of amorphous ice would release enough heat to sublimate the surrounding crystalline ice in a volume equivalent to the observed pits.



**Sinkhole model.** A first order estimate of the stability of a cavity ceiling may be derived by treating the ceiling a beam failing under its own weight<sup>18,38</sup>. Failure of this beam occurs when the bending moment exceeds the material's tensile strength. Assuming the comet's material is highly porous, the stable beam depth  $d = 6D^2\rho a/(8S)$ , where  $D$  is the cavity's diameter,  $\rho$  is the density of the material in the ceiling (assumed to be  $470 \text{ kg m}^{-3}$ ),  $a$  is the gravitational acceleration on the comet<sup>12</sup> ( $5 \times 10^{-4} \text{ m s}^{-2}$ ) and  $S$  is the tensile strength of the ceiling material. For the tensile strength, we adopted an initial range between the lower limit of 10 Pa derived from overhangs on the surface<sup>19</sup> and the upper limit of 10 kPa derived from the Deep Impact experiment<sup>39</sup>. We further assume that the cavity is of approximately the same size as the resulting pit and that the depth of the pit is comparable to the depth of the original ceiling.

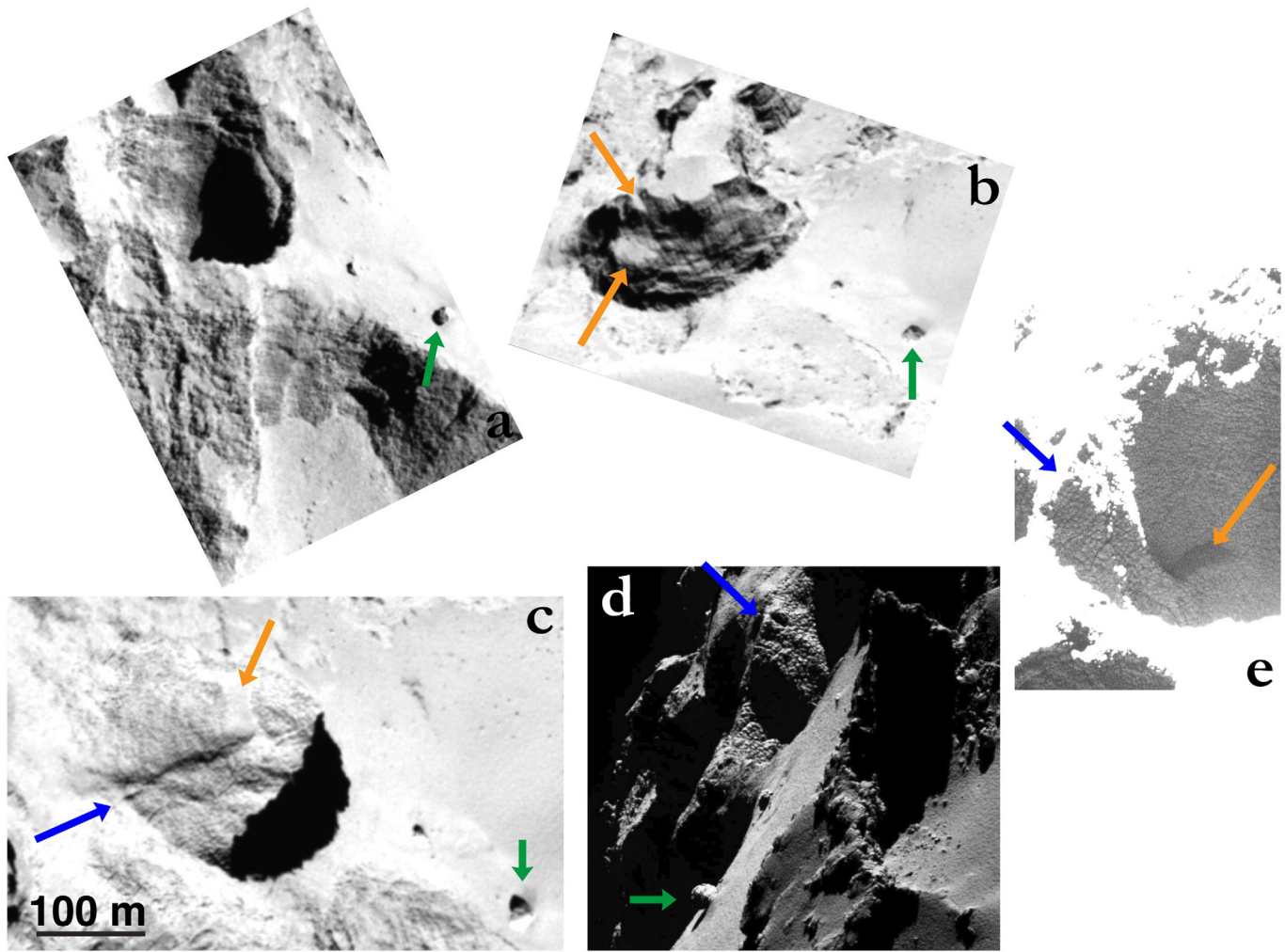
**Code availability.** The code used to generate the orbital evolution of comet 67P is a direct implementation of a published model<sup>32</sup>.

28. Delsemme, A. H. in *Comets* (ed. Wilkening, L. L.) 85–130 (Univ. Arizona Press, 1982).
29. Sekanina, Z. Anisotropic emission from comets: fans versus jets. I. Concept and modeling. In *Proc. International Symposium on the Diversity and Similarity of Comets*, Vol. ESA SP-278, 315–322 (1987).
30. Belton, M. J. S. Cometary activity, active areas, and a mechanism for collimated outflows on 1P, 9P, 19P, and 81P. *Icarus* **210**, 881–897 (2010).
31. Crifo, J.-F., Rodionov, A. V., Szegő, K. & Fulle, M. Challenging a paradigm: do we need active and inactive areas to account for near-nuclear jet activity? *Earth Moon Planets* **90**, 227–238 (2002).
32. Marsden, B. G., Sekanina, Z. & Yeomans, D. K. Comets and non-gravitational forces. *V. Astron. J.* **78**, 211–225 (1973).
33. Tozzi, G. P. *et al.* Evolution of the dust coma in comet 67P/Churyumov-Gerasimenko before the 2009 perihelion. *Astron. Astrophys.* **531**, A54 (2011).
34. Lara, L. M., Lin, Z.-Y., Rodrigo, R. & Ip, W.-H. 67P/Churyumov-Gerasimenko activity evolution during its last perihelion before the Rosetta encounter. *Astron. Astrophys.* **525**, A36 (2011).
35. Snodgrass, C. *et al.* Beginning of activity in 67P/Churyumov-Gerasimenko and predictions for 2014–2015. *Astron. Astrophys.* **557**, A33 (2013).
36. Ghormley, J. A. Enthalpy changes and heat-capacity changes in the transformations from high-surface-area amorphous ice to stable hexagonal ice. *J. Chem. Phys.* **48**, 503–508 (1968).
37. De Sanctis, M. C., Lasue, J. & Capria, M. T. Seasonal effects on comet nuclei evolution: activity, internal structure, and dust mantle formation. *Astron. J.* **140**, 1–13 (2010).
38. White, W. B. *Geomorphology and Hydrology of Karst Terrains* (Oxford Univ. Press, 1988).
39. Richardson, J. E. & Melosh, H. J. An examination of the Deep Impact collision site on comet Tempel 1 via Stardust-NExT: placing further constraints on cometary surface properties. *Icarus* **222**, 492–501 (2013).



**Extended Data Figure 1 | Perihelion distance of comet 67P as a function of time.** Solid line, mean value of the orbits integrated according to a Monte Carlo method. Dashed lines, standard deviation of the mean value. **a**, Perihelion

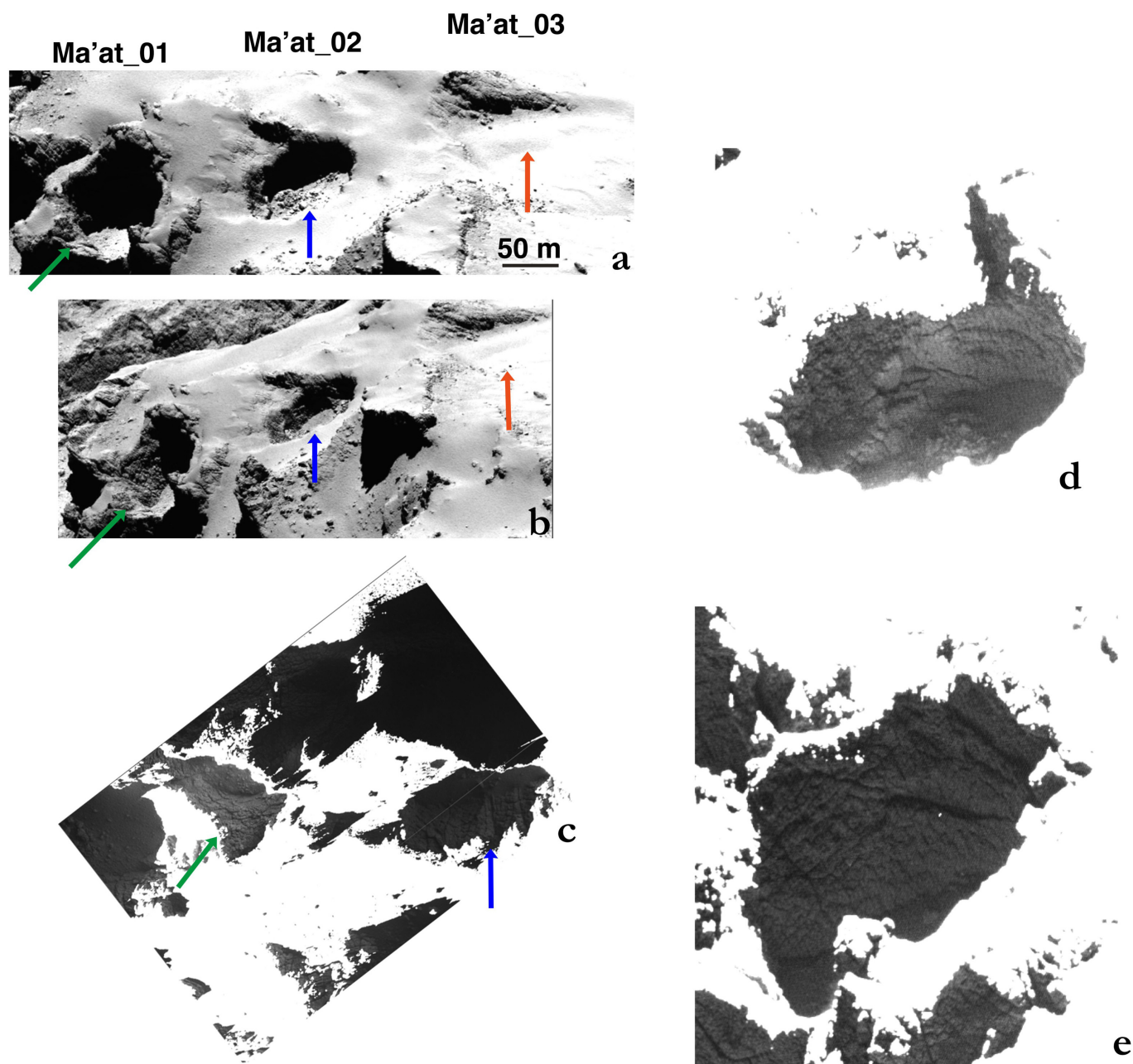
distance over the last 270 years, when comet 67P experienced several close encounters with Jupiter. **b**, The long term integration over the full dynamical lifetime of the comet (10,000 years).



**Extended Data Figure 2 | Multiple views of the Seth\_01 pit observed by the OSIRIS camera.** **a**, Southern part of the pit wall; **b**, western part of the pit wall; **c**, **d**, eastern part of the pit wall with different illumination conditions; and **e**, southeastern part of the pit wall observed in the shadow. In all the images, the

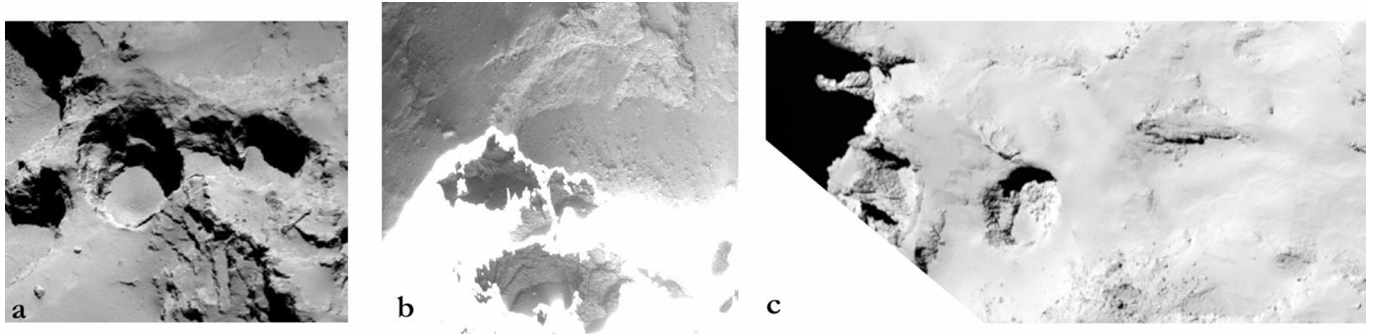
green arrow points to the same boulder and the blue arrow to the same ridge inside the pit. The orange arrows point to terraces within the pit. The Seth\_01 pit is 220 m in diameter.





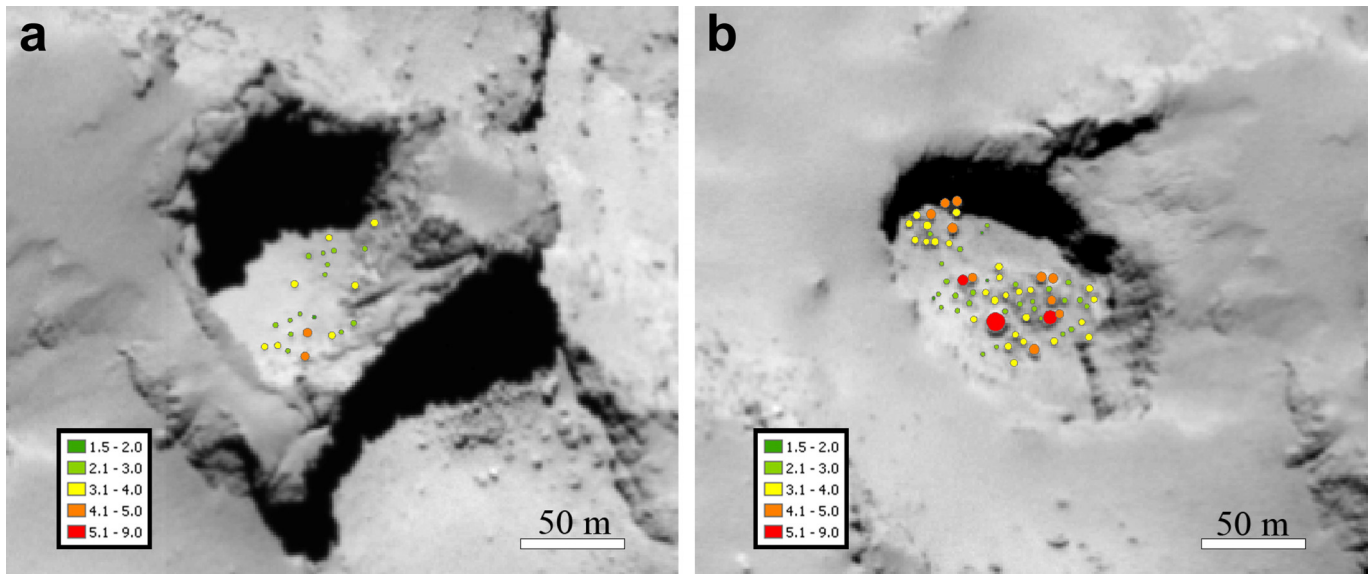
**Extended Data Figure 3 | Multiple views of the Ma'at\_01, Ma'at\_02 and Ma'at\_03 pits observed by the OSIRIS camera.** a, b, Side views of the pits with different illumination conditions; c, opposite viewing conditions highlighting the other side in the shadow; and d, e, detailed views of Ma'at\_02 (d) and Ma'at\_01 (e) from light reflection in the shadow. Note the clear cross-cutting

fractures on the wall in e. In c, the white line is an artefact due to stretching of the image to highlight the shadowed part. The Ma'at\_02 pit is 130 m in diameter. The blue, green and orange arrows point to the same features in each image.



**Extended Data Figure 4 | Additional views of the Seth\_01 and Ma'at\_01 pits.** **a**, The floor of Seth\_01 shows no accumulation of boulders; the same is true for Seth\_02 and Seth\_03 (not shown). **b**, The floor of Ma'at\_01 shows a few

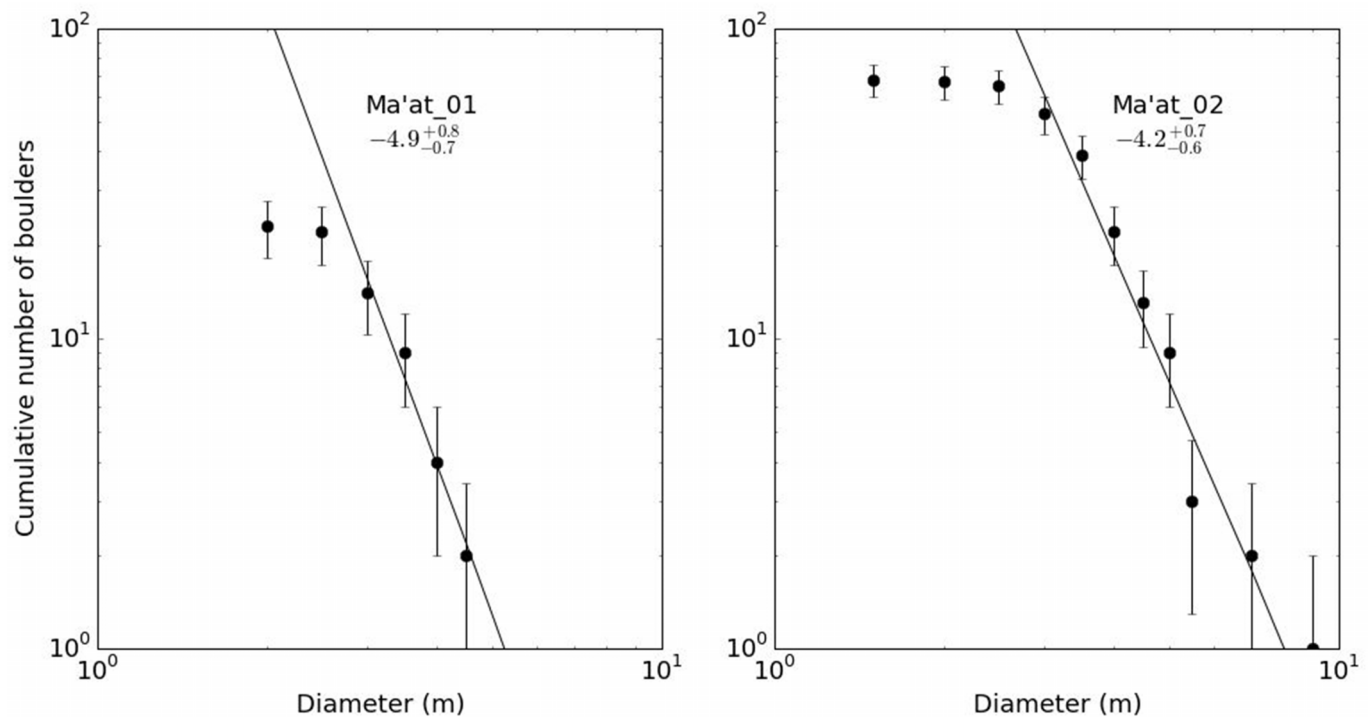
boulders that have accumulated; note the activity located at the bottom. **c**, The floor of Ma'at\_02 shows an asymmetric accumulation of boulders that could be the result of upper wall collapse.



**Extended Data Figure 5 | Boulder counts in Ma'at\_01 and Ma'at\_02.** We counted boulders on the floor of Ma'at\_01 and Ma'at\_02. We used OSIRIS narrow angle camera (NAC) images with a resolution of 1.2 metres per pixel, acquired at 67 km from the comet nucleus centre. **a, b,** The illumination conditions are such that almost 80% of the floor of Ma'at\_01 (**a**) and 95% of the floor of Ma'at\_02 (**b**) are illuminated and the pits are facing the observer, which ensures an unbiased boulder count. We identified 23 boulders inside

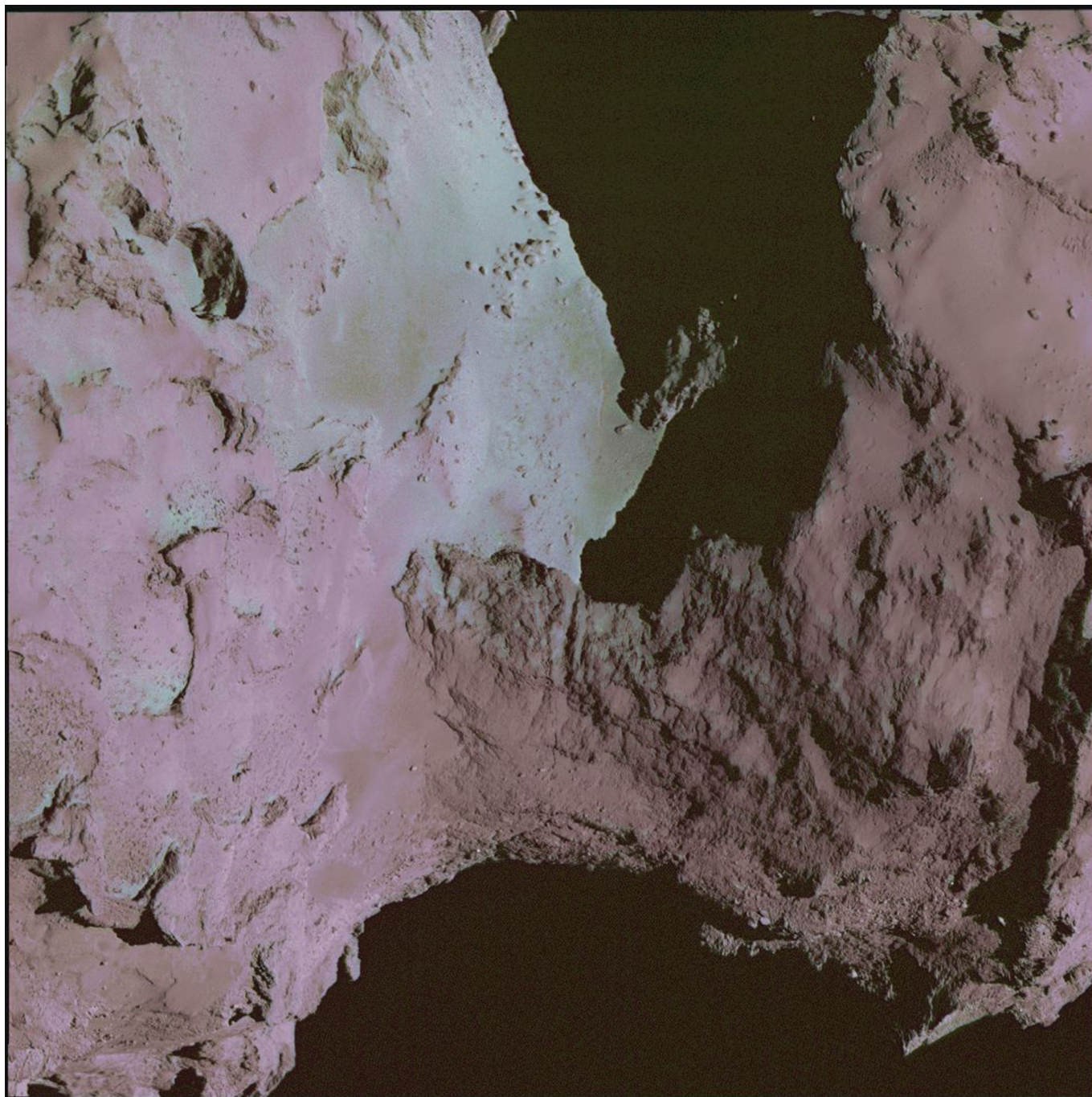
Ma'at\_01 and 68 on the floor of Ma'at\_02. The diameter of the boulders (in metres) is indicated by the coloured circles; see inset. Despite the 1.2 metres per pixel resolution, we were able to identify some boulders with a diameter between 1.5 m and 2.5 m (9 in Ma'at\_01 and 15 in Ma'at\_02), owing to the presence of elongated shadows. The maximum boulder diameter is 4.3 m in Ma'at\_01 and 9.0 m in Ma'at\_02.





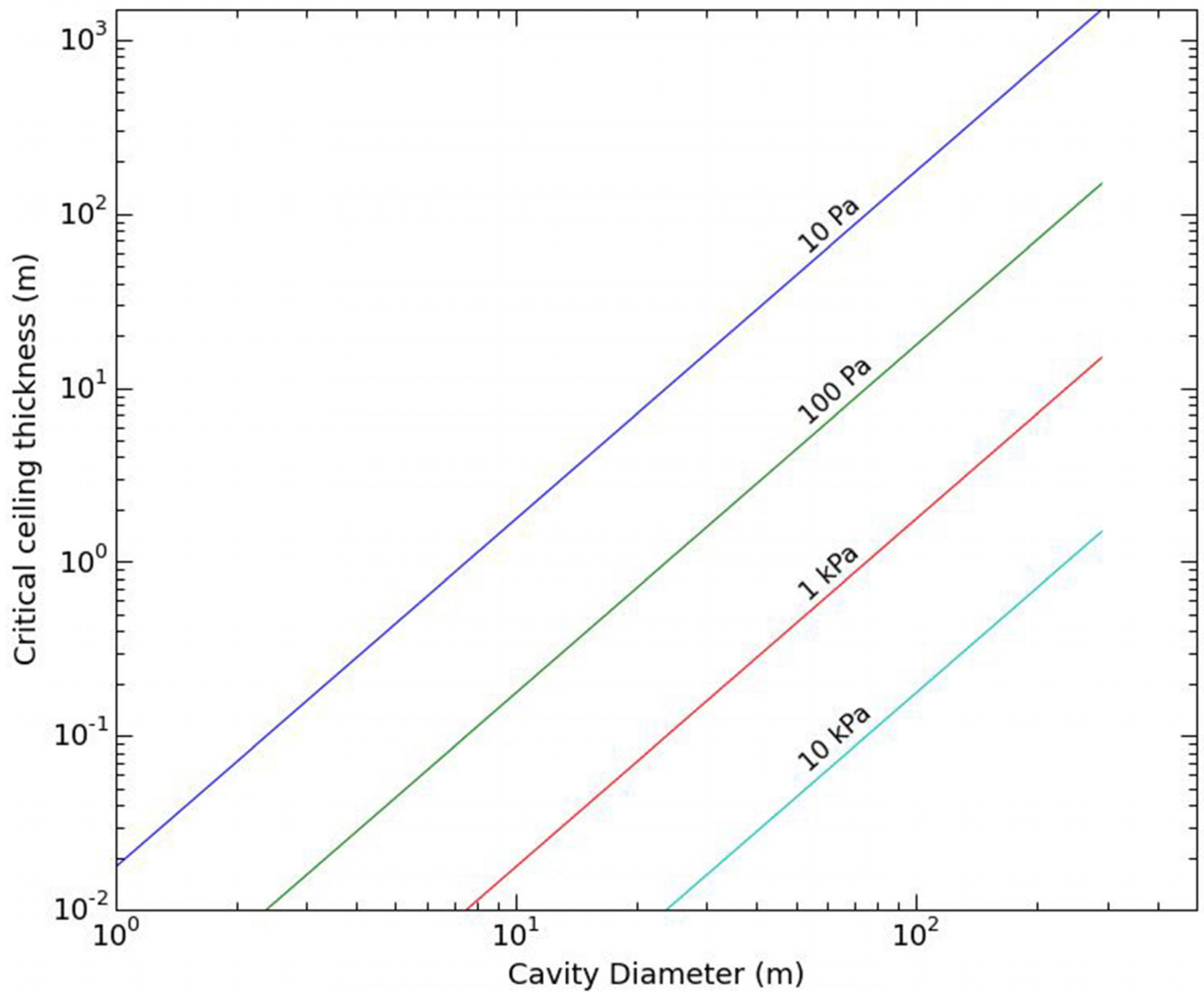
**Extended Data Figure 6 | Cumulative boulder-size distribution for Ma'at\_01 and Ma'at\_02.** This distribution has a power index of  $-4.9^{+0.8}_{-0.7}$  for Ma'at\_01 (left) and  $-4.2^{+0.7}_{-0.6}$  for Ma'at\_02 (right), for boulder diameters greater than 3 m; the corresponding power laws are indicated by the solid (fit) lines. Boulders smaller than 3 m in diameter are at the edge of our detection limit, meaning that the counts for these boulders are less reliable than the other

counts; consequently, they were not included when fitting the power law. The higher number of boulders in Ma'at\_02 is consistent with the theory that boulders are debris that falls from the walls as the pit erodes, long after the initial formation of the pit. Error bars are defined as the square root of the cumulative number of boulders to reflect the increasing diameter uncertainty for small boulder sizes.



**Extended Data Figure 7 | RGB view of the Seth pits and the Hapi region.** The red–blue–green components of this colour map represent colour ratios between the reflectance signals measured at different wavelengths: red, 989 nm/649 nm; green, 480 nm/649 nm; blue, 649 nm. The colour map is overlaid

onto a grey image showing the comet surface. The Hapi region and part of Seth appear with a blue hue, indicative of a bluer spectral slope than other regions of the nucleus, which are typically red. The interior of Seth\_01, Seth\_02 and Seth\_03 have the same blue hue that is characteristic of the active Hapi region.



**Extended Data Figure 8 | Modelled critical ceiling thickness for increasing cavity diameter and different tensile strengths.** We predict the average tensile strength of a collapsed layer using the dimensions of a pit (Methods). For

example, a pit of 220 m in diameter and 185 m in depth (such as, Seth\_01) suggests that the collapsed layer had an average tensile strength of 50 Pa.



Extended Data Table 1 | List of pits considered in this paper

| Pit Id    | Latitude (°) | Longitude (°) | Diameter (m) | Depth (m) | d/D  | Jets detected? |
|-----------|--------------|---------------|--------------|-----------|------|----------------|
| Seth_01   | 70           | 220           | 220          | 185       | 0.84 | Yes            |
| Seth_02   | 69           | 205           | 110          | 95        | 0.87 | Yes            |
| Seth_03   | 68           | 195           | 140          | 100       | 0.72 | Yes            |
| Seth_04   | 78           | 198           | 160          | 130       | 0.77 | Yes            |
| Seth_05   | 57           | 239           | 230          | 210       | 0.92 | Yes            |
| Seth_06   | 39           | 204           | 210          | 50        | 0.24 | No             |
| Ash_01    | 30           | 90            | 70           | 10        | 0.14 | No             |
| Ash_02    | 36           | 85            | 50           | 13        | 0.26 | No             |
| Ash_03    | 23           | 204           | 126          | 42        | 0.33 | No             |
| Ash_04    | 20           | 210           | 192          | 21        | 0.11 | No             |
| Ash_05    | 20           | 216           | 300          | 64        | 0.21 | No             |
| Ash_06    | 23           | 226           | 310          | 87        | 0.28 | No             |
| Ma'at_01  | 42           | 8             | 125          | 65        | 0.52 | Yes            |
| Ma'at_02  | 35           | 9             | 130          | 60        | 0.47 | Yes            |
| Ma'at_03  | 35           | 2             | 140          | 50        | 0.37 | No             |
| Ma'at_04  | -10          | 327           | 115          | 20        | 0.18 | No             |
| Bastet_01 | 3            | 21            | 80           | 38        | 0.48 | No             |
| Hathor_01 | 25           | 26            | 53           | 21        | 0.40 | No             |

Diameter and depth measured on digital terrain model reconstructed from OSIRIS images by stereo-photogrammetry<sup>12</sup>. Active pits have a mean  $d/D = 0.73 \pm 0.08$ ; inactive pits have a mean  $d/D = 0.26 \pm 0.08$ . Maximum error is 20 m for the diameter and 5 m for the depth. Coordinates are given in the 'Cheops' reference frame<sup>12</sup>.

Extended Data Table 2 | List of images used

| <b>Pits in Seth region</b>                       | <b>Distance from nucleus center (km)</b> | <b>Resolution on the surface (m/px)</b> |
|--|--|---|
| NAC_2014-08-06T02.19.14.570Z_ID30_1397549900_F22 | 117.2                                    | 2.14                                    |
| NAC_2014-08-21T11.42.53.641Z_ID30_1397549800_F22 | 70.6                                     | 1.28                                    |
| NAC_2014-08-21T13.42.54.561Z_ID30_1397549200_F22 | 70.95                                    | 1.28                                    |
| NAC_2014-08-28T20.42.53.590Z_ID30_1397549900_F22 | 54.24                                    | 0.97                                    |
| NAC_2014-08-29T04.42.56.583Z_ID30_1397549200_F22 | 52.90                                    | 0.98                                    |
| NAC_2014-08-29T20.42.53.538Z_ID00_1397549900_F22 | 54.06                                    | 0.97                                    |
| NAC_2014-09-02T21.44.22.575Z_ID30_1397549800_F22 | 62.16                                    | 1.12                                    |
| NAC_2014-09-20T01.48.48.382Z_ID30_1397549200_F22 | 26.2                                     | 0.45                                    |
| NAC_2014-10-02T00.26.22.560Z_ID30_1397549300_F22 | 18.65                                    | 0.31                                    |
| NAC_2014-10-02T12.11.26.599Z_ID30_1397549700_F22 | 18.86                                    | 0.31                                    |
| NAC_2014-10-03T13.40.22.589Z_ID30_1397549900_F22 | 19.18                                    | 0.32                                    |
| WAC_2014-10-20T08.15.50.752Z_ID30_1397549000_F18 | 7.2                                      | 0.54                                    |

| <b>Pits in Ma'at region</b>                      | <b>Distance from nucleus center (km)</b> | <b>Resolution on the surface (m/px)</b> |
|--|--|---|
| NAC_2014-08-21T20.42.54.581Z_ID30_1397549100_F22 | 67.04                                    | 1.21                                    |
| NAC_2014-08-21T21.42.54.553Z_ID30_1397549300_F22 | 66.54                                    | 1.20                                    |
| NAC_2014-08-22T21.41.54.592Z_ID00_1397549000_F22 | 60.14                                    | 1.08                                    |
| NAC_2014-09-12T01.33.04.375Z_ID30_1397549400_F22 | 27.55                                    | 0.48                                    |
| WAC_2014-09-12T04.10.28.751Z_ID30_1397549100_F17 | 27.57                                    | 2.63                                    |
| NAC_2014-09-13T03.36.12.463Z_ID30_1397549200_F22 | 30.05                                    | 0.52                                    |
| NAC_2014-10-03T17.37.22.649Z_ID30_1397549100_F22 | 19.19                                    | 0.32                                    |
| NAC_2014-10-04T05.27.21.576Z_ID30_1397549100_F22 | 19.19                                    | 0.32                                    |
| NAC_2014-10-14T21.40.03.314Z_ID30_1397549700_F22 | 10.57                                    | 0.16                                    |
| WAC_2014-10-20T04.52.10.460Z_ID10_1397549600_F18 | 9.31                                     | 0.75                                    |

Also shown is the distance from the comet nucleus centre that the image was taken and the resolution of the image.

# A colloidal quantum dot spectrometer

Jie Bao<sup>1,2,3</sup> & Mouni G. Bawendi<sup>2</sup>

Spectroscopy is carried out in almost every field of science, whenever light interacts with matter<sup>1</sup>. Although sophisticated instruments with impressive performance characteristics are available, much effort continues to be invested in the development of miniaturized, cheap and easy-to-use systems<sup>1–13</sup>. Current microspectrometer designs mostly use interference filters<sup>2–5</sup> and interferometric optics<sup>3</sup> that limit their photon efficiency, resolution and spectral range<sup>2,3</sup>. Here we show that many of these limitations can be overcome by replacing interferometric optics with a two-dimensional absorptive filter array composed of colloidal quantum dots<sup>14–17</sup>. Instead of measuring different bands of a spectrum individually after introducing temporal or spatial separations with gratings or interference-based narrow-band filters, a colloidal quantum dot spectrometer measures a light spectrum based on the wavelength multiplexing principle<sup>18</sup>: multiple spectral bands are encoded and detected simultaneously with one filter and one detector<sup>9–12</sup>, respectively, with the array format allowing the process to be efficiently repeated many times using different filters with different encoding so that sufficient information is obtained to enable computational reconstruction of the target spectrum. We illustrate the performance of such a quantum dot microspectrometer, made from 195 different types of quantum dots with absorption features that cover a spectral range of 300 nanometres, by measuring shifts in spectral peak positions as small as one nanometre. Given this performance, demonstrable avenues for further improvement, the ease with which quantum dots can be processed and integrated, and their numerous finely tuneable bandgaps that cover a broad spectral range, we expect that quantum dot microspectrometers will be useful in applications where minimizing size, weight, cost and complexity of the spectrometer are critical.

Figure 1 illustrates different operating mechanisms that enable spectral analysis. Conventionally, spectral analysis is achieved using dispersive optics such as gratings, which separate different light bands before their intensity is measured. An alternative approach uses a set of distinct, yet continuously tuneable, broadband absorptive filters and reconstructs the target spectrum from the set of measured broadband intensity distributions. In principle, this approach allows for a simplified and cheaper spectrometer design that is well suited to miniaturization. But suitable broadband filters, with appropriate system compatibility and photo-stability, have so far not been readily available. Here we show that colloidal quantum dots (CQDs), with their continuously tuneable, size-dependent bandgaps, offer a practical solution to the filter availability problem that has hindered the development of microspectrometers that use broadband absorptive filters.

CQDs are semiconductor nanocrystals with radii smaller than the bulk exciton Bohr radius, which leads to quantum confinement of electronic charges, with decreasing crystal size strengthening confinement and hence increasing the effective bandgap, and blue-shifting both optical absorption and fluorescent emission. Research over the past few decades has resulted in a vast library of CQDs. We can now tune absorption spectra continuously and finely over wavelengths ranging from deep ultraviolet to mid-infrared, simply by changing the CQD's size, shape and composition<sup>19</sup>. CQDs can also be printed

into very fine patterns<sup>20,21</sup>. They have superior photo-stability than dyes<sup>22,23</sup>. In addition, dyes achieve different absorption profiles by altering their chemical compositions and structures, which leads to them having different chemical properties<sup>24</sup>. In contrast, CQDs can provide different bandgaps but still share similar chemical properties, which simplifies synthesis and processing when a large number of materials are involved. CQDs are thus ideal broadband filter materials for use in wavelength multiplexing microspectrometers.

Figure 2a illustrates how a basic quantum dot spectrometer operates. The measurement of an incident-light spectrum by a quantum dot spectrometer is based on the measurement of the total transmitted intensity of the spectrum that passes through a given CQD filter. This intensity measurement is repeated for each CQD filter, resulting in a set of measured transmitted-light intensities. The original (incident) light spectrum is computationally reconstructed on the basis of this set of transmitted-light intensities. Spectral reconstruction is performed using linear regression, by finding a spectrum (the reconstructed spectrum) that minimizes the sum of the squares of the differences between the set of measured transmitted-light intensities and a set of numerically computed intensities, which are the products of the reconstructed incident-light spectrum and the transmission spectra that are measured for each CQD filter, integrated with respect to wavelength.

To illustrate the above method mathematically, consider an arbitrary incident-light spectrum  $\Phi(\lambda)$  (where  $\lambda$  is the wavelength) that is transmitted through a set of  $n_F$  CQD filters, one at a time, resulting in measured total transmitted-light intensities  $I_i$  (where  $i = 1, 2, \dots, n_F$  is the filter number). One data point is generated per filter:

$$I_i = \sum_{\lambda} \Phi(\lambda) T_i(\lambda), \quad i = 1, 2, \dots, n_F \quad (1)$$

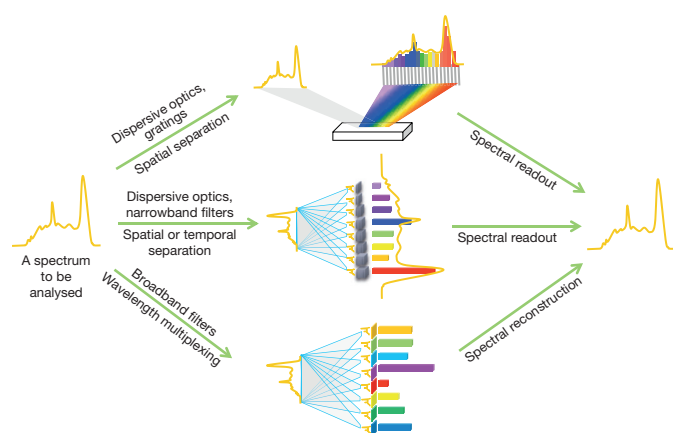
where  $T_i(\lambda)$  is the transmission spectrum of CQD filter  $i$ . Because  $T_i(\lambda)$  for each CQD filter is predetermined (by measuring it), the only unknown is  $\Phi(\lambda)$ , which is evaluated at  $n_\lambda$  discrete  $\lambda$  values and so corresponds to a set of  $n_\lambda$  variables. Although increasing  $n_\lambda$  for a given spectral range may increase the spectral resolution, being able to solve equation (1) limits  $n_\lambda$  to be at most equal to  $n_F$ . In the ideal case,  $n_\lambda = n_F$ , which produces a set of linear equations with a unique solution; in practice, however, measurement errors yield inconsistencies within the set of equations. Approximate solutions can nevertheless be obtained using least-squares linear regression. In such a case, a given  $n_F$  no longer provides an equal number of accurately reconstructed spectral data points and so  $n_\lambda < n_F$ . The larger the error, the more filters are required to accurately reconstruct each spectral data point.

The operation of a quantum dot spectrometer can be made more efficient by measuring the set of intensities simultaneously with a collection of light detectors, each of which is coupled to a designated CQD filter (Fig. 2b). If the collection of light detectors is arranged into a miniaturized, two-dimensional detector array such as a charge-coupled device (CCD) array detector and the set of CQD filters is integrated into a CQD filter array and coupled to the detector array, then a quantum dot spectrometer can perform spectral measurements in a snapshot fashion, without scanning or switching filters.

<sup>1</sup>Department of Electronic Engineering, Tsinghua University, Beijing, 100084, China. <sup>2</sup>Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

<sup>3</sup>Department of Physics, California Institute of Technology, Pasadena, California 91125, USA.

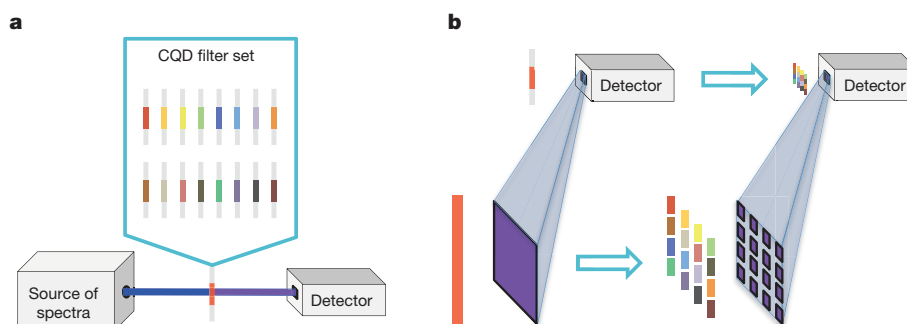




**Figure 1 | Comparison of spectrometer mechanisms.** Using a grating-based spectrometer (top path), different bands of a light spectrum are first separated. Then the intensity of each band is measured individually. Using a narrowband-filter-based spectrometer (middle path), only one band of a light spectrum passes through one filter at a time, which results in the detection of the intensity of one band. Different bands of a spectrum are detected with different filters. These filters and detectors may be either spatially or temporally separated from each other; a set of spatially separated, discrete interference filters is shown here. Using a broadband-filter-based spectrometer (bottom path), multiple spectral bands pass through each broadband filter simultaneously, reaching a single detector and resulting in the detection of the intensity of multiple bands. This process is repeated with multiple broadband filters, each of which modifies the spectrum in a different way. The original spectrum is computationally reconstructed from the intensities detected from each broadband filter. This process is known as wavelength multiplexing. The quantum dot spectrometer presented here is a broadband-filter-based spectrometer that uses CQDs as broadband filter material.

We demonstrate a quantum dot microspectrometer by fabricating one that operates over a spectral range of 300 nm (390–690 nm), and uses 195 different CQD filters (Fig. 3a, Methods subsection ‘Additional experimental information’). Each filter is unique, as a result of tuning the quantum dot size or composition (cadmium sulfide, CdS, or cadmium selenide, CdSe), with the effective bandgaps spread roughly evenly over the selected spectral range (Fig. 3b, Extended Data Fig. 1). Figure 3c shows a photograph of a quantum dot spectrometer formed by coupling a CQD filter array to a CCD, which was integrated into a digital camera (Methods subsection ‘Additional experimental information’, Extended Data Fig. 2).

To illustrate the performance of the quantum dot spectrometer, Fig. 4a–d presents measurements of four arbitrary spectra (but with varying intensity levels and widths) that were generated using a white-light source and a set of optical filters. The quantum dot spectrometer was able to reproduce all the major features of the spectra.



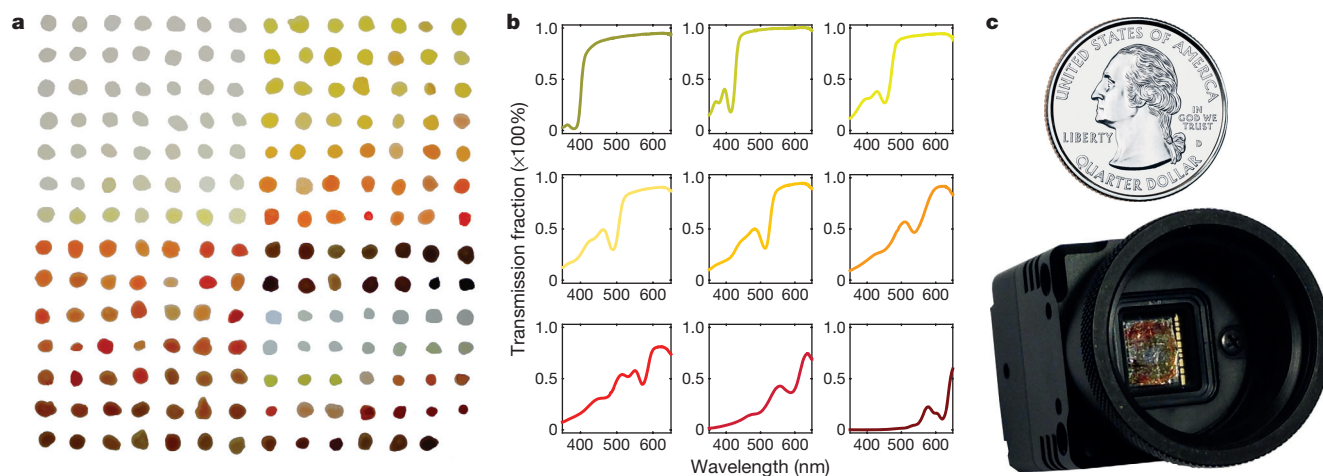
**Figure 2 | Operation of quantum dot spectrometers.** **a**, A basic quantum dot spectrometer is composed of a set of CQD absorptive filters and a light detector. A spectrum is measured by placing the CQD filters in front of the detector one at a time and measuring the light intensities passing through each. **b**, Instead of measuring the light intensities using one detector and one filter at

Deviations between the reconstructed spectra and the reference spectra measured using a commercial spectrometer (Methods subsection ‘Additional experimental information’) arise largely from subtle features, which are due to system measurement errors and the number of CQD filters used. The quantum dot spectrometer was also used to measure the emission spectra of five fluorescent CQD samples (Fig. 4e) with emission peaks ranging from 450 nm to 650 nm. As shown in Fig. 4f, the emission spectra measured by the quantum dot spectrometer match those measured with a standard spectrofluorometer (Methods subsection ‘Additional experimental information’). To demonstrate the spectral resolution of the quantum dot spectrometer, we used monochromatic light with a bandwidth of 2 nm, centred at different wavelengths, and obtained accurate spectra (Fig. 4g); even wavelength shifts in the peak positions as small as 1 nm are accurately measured (Fig. 4h). Further analysis shows that the quantum dot spectrometer resolves two peaks separated by 2–3 nm (Methods subsection ‘Characterization of spectral resolving capability’, Extended Data Fig. 3).

Spectral resolution and spectral reconstruction accuracy can be improved by reducing measurement errors, for example, by using a more stable calibration light source and a higher-performance CCD detector (Methods subsection ‘Effects of measurement errors on spectral resolving capability’, Extended Data Figs 4–6). Spectral resolution and spectral range can be improved by using a larger number of CQDs, which cover a broader spectral range (Methods subsection ‘Effects of the number of filters on spectral resolving capability’, Extended Data Fig. 7). Another strategy for improving spectral resolution and reconstruction accuracy is to use prior knowledge of the spectra (such as the sparseness of the signals<sup>25</sup>) and more sophisticated reconstruction algorithms such as  $L^1$ -norm minimization<sup>26</sup>, principal component regression<sup>27</sup>, and genetic algorithms<sup>28</sup> (Methods subsection ‘Effects of algorithms on spectral reconstruction accuracy’, Extended Data Fig. 8). Our simulations have also shown that the dynamic range can be increased by reducing measurement errors (Methods subsection ‘Effects of measurement errors on dynamic range’, Extended Data Table 1). Finally, recent work on air-stable quantum dot solar cells<sup>29</sup> as well as the present experiments indicate that these quantum dot devices are sufficiently stable for long-term practical use (Methods subsection ‘Stability analysis’, Extended Data Fig. 9).

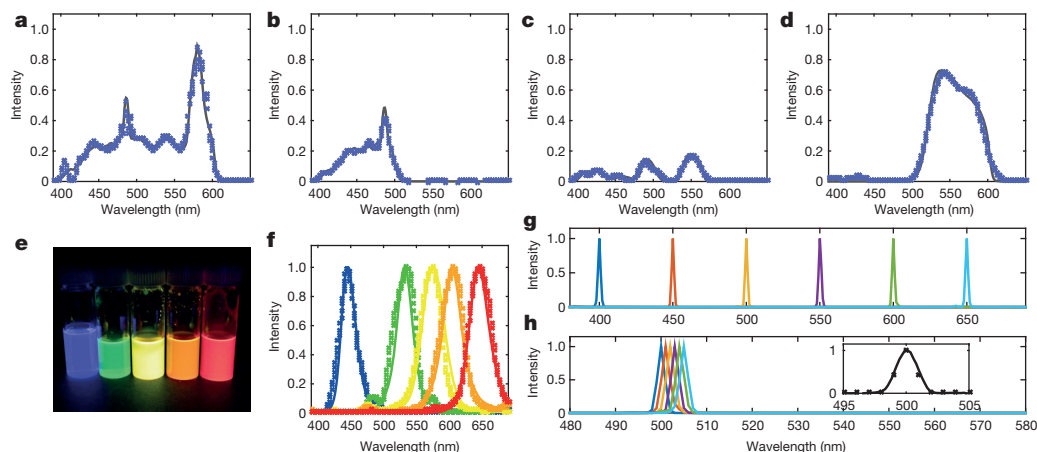
Quantum dot spectrometers are promising high-performance microspectrometers because spectral resolution and spectral range can be increased simultaneously, simply by increasing the number of different CQDs used in the filter array. The wavelength multiplexing approach means that these improvements are achieved without sacrificing total photon efficiency. Another advantage of the quantum dot spectrometer is its ability to analyse light from a source with a wide angular distribution while maintaining the spectral resolution, as demonstrated by the use of an integrating sphere between the light source and the spectrometer (Methods subsection ‘Additional experimental information’); this

at a time, a more efficient quantum dot spectrometer (such as the one demonstrated here and illustrated in Fig. 3c) measures the set of intensities in parallel by using an array detector, with each detecting element dedicated to one CQD filter, all of which are integrated into a CQD filter array.



**Figure 3 | CQD filters and an integrated quantum dot spectrometer.** **a**, 195 CQD materials in the form of filters. Each dot is a CQD filter made of one type of CQD material embedded in a polyvinyl butyral thin film. **b**, Transmission spectra for some of the CQD filters shown in **a**. Transmission spectra of all 195

CQD filters can be found in Extended Data Fig. 1. **c**, A quantum dot microspectrometer in the form of a digital camera with electronics and circuits, made from the 195 CQD filters and a CCD array detector, which is comparable in size to a US quarter.



**Figure 4 | Quantum dot spectrometer measurements.** **a–d**, Crosses correspond to the quantum dot spectrometer measurement of broadband spectra; solid lines correspond to reference spectra (measured using a commercial spectrometer). **e**, Fluorescent emission of five CQD samples under ultraviolet excitation. **f**, Markers correspond to measurements (using the quantum dot spectrometer) of the emission spectra of the five CQD samples shown in **e**; solid lines correspond to reference spectra (measured using a

spectrofluorometer). **g**, **h**, Measurements of monochromatic light spectra. The peak positions of the six monochromatic lights in **g** are 400 nm, 450 nm, 500 nm, 550 nm, 600 nm and 650 nm. The peak positions of the six monochromatic lights in **h** are 500 nm, 501 nm, 502 nm, 503 nm, 504 nm and 505 nm. The inset of **h** compares the measured spectrum of 500-nm monochromatic light (crosses) and the reference spectrum (solid line).

means it has a larger etendue than interferometric-optics-based spectrometers. Lastly, simplifying the system by replacing dispersive optics with CQD filters allow for a notable size reduction of the device. Given the ability to make CCD pixels less than 10  $\mu\text{m}$  in size and printed CQD features<sup>20,21</sup>, we expect that it is possible to produce quantum dot spectrometers that are orders of magnitude smaller than our demonstration system, without compromising spectral range or resolution.

Previous applications of CQDs largely relied on generating materials (CQDs) with a high fluorescence quantum yield (ratio of the number of photons emitted to that absorbed), but such materials are difficult to synthesize and maintain. However, high fluorescence quantum yield is not needed in a quantum dot spectrometer, which is based on the most reliable property of CQDs—their light absorption characteristics. This makes it feasible to replace the CQD materials used in the quantum dot spectrometer demonstrated here with more environmentally friendly, but relatively lower quantum yield, cadmium-free materials such as zinc selenide (ZnSe) and indium phos-

phide (InP)<sup>19</sup>, which adds to the appeal of CQDs as a step towards producing single-chip high-performance spectrometers that are smaller than a mobile phone camera image sensor. Such spectrometers may be of benefit to space exploration, in surgical and clinical settings, to personalized health care and to lab-on-a-chip diagnostics, for example, wherever minimizing the size, weight, cost or complexity of spectrometers is critical<sup>1,30</sup>.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 15 August 2013; accepted 12 May 2015.

1. Bacon, C. P., Mattley, Y. & DeFrece, R. Miniature spectroscopic instrumentation: applications to biology and chemistry. *Rev. Sci. Instrum.* **75**, 1–16 (2004).
2. Wang, S. W. *et al.* Concept of a high-resolution miniature spectrometer using an integrated filter array. *Opt. Lett.* **32**, 632–634 (2007).
3. Wolffenbuttel, R. F. State-of-the-art in integrated optical microspectrometers. *IEEE Trans. Instrum. Meas.* **53**, 197–202 (2004).

4. Emadi, A., Wu, H., de Graaf, G. & Wolffenbuttel, R. Design and implementation of a sub-nm resolution microspectrometer based on a linear-variable optical filter. *Opt. Express* **20**, 489–507 (2012).
5. Demro, J. C. *et al.* Design of a multispectral, wedge filter, remote-sensing instrument incorporating a multiport, thinned, CCD area array. *Proc. SPIE* **2480**, 280–286 (1995).
6. Gan, X., Pervez, N., Kyriassis, I., Hatami, F. & Englund, D. A high-resolution spectrometer based on a compact planar two dimensional photonic crystal cavity array. *Appl. Phys. Lett.* **100**, 231104 (2012).
7. Laux, E., Genet, C., Skauli, T. & Ebbesen, T. W. Plasmonic photon sorters for spectral and polarimetric imaging. *Nature Photon.* **2**, 161–164 (2008).
8. Xu, T., Wu, Y.-K., Luo, X. & Guo, L. J. Plasmonic nanoresonators for high-resolution colour filtering and spectral imaging. *Nature Commun.* **1**, 59 (2010).
9. Knipp, D. *et al.* Silicon-based micro-fourier spectrometer. *IEEE Trans. Electron. Dev.* **52**, 419–426 (2005).
10. le Coarer, E. *et al.* Wavelength-scale stationary-wave integrated Fourier-transform spectrometry. *Nature Photon.* **1**, 473–478 (2007).
11. Pervez, N. K. *et al.* Photonic crystal spectrometer. *Opt. Express* **18**, 8277–8285 (2010).
12. Redding, B., Liew, S. F., Sarma, R. & Cao, H. Compact spectrometer based on a disordered photonic chip. *Nature Photon.* **7**, 746–751 (2013).
13. Garini, Y., Young, I. T. & McNamara, G. Spectral imaging: principles and applications. *Cytometry A* **69A**, 735–747 (2006).
14. Brus, L. E. A simple model for the ionization potential, electron affinity, and aqueous redox potentials of small semiconductor crystallites. *J. Chem. Phys.* **79**, 5566–5571 (1983).
15. Efros, A. L. & Efros, A. L. Interband absorption of light in a semiconductor sphere. *Sov. Phys. Semicond.* **16**, 772–775 (1982).
16. Murray, C. B., Norris, D. J. & Bawendi, M. G. Synthesis and characterization of nearly monodisperse CdE (E = sulfur, selenium, tellurium) semiconductor nanocrystallites. *J. Am. Chem. Soc.* **115**, 8706–8715 (1993).
17. Alivisatos, A. P. Semiconductor clusters, nanocrystals, and quantum dots. *Science* **271**, 933–937 (1996).
18. James, J. F. & Sternberg, R. S. in *The Design of Optical Spectrometers* Ch. 8 (Chapman & Hall, 1969).
19. Rogach, A. (ed.) *Semiconductor Nanocrystal Quantum Dots: Synthesis, Assembly, Spectroscopy and Applications* 1–72 (Springer, 2008).
20. Kim, L. *et al.* Contact printing of quantum dot light-emitting devices. *Nano Lett.* **8**, 4513–4517 (2008).
21. Kim, T. *et al.* Full-colour quantum dot displays fabricated by transfer printing. *Nature Photon.* **5**, 176–182 (2011).
22. Giepmans, B. N. G., Adams, S. R., Ellisman, M. H. & Tsien, R. Y. The fluorescent toolbox for assessing protein location and function. *Science* **312**, 217–224 (2006).
23. Resch-Genger, U. *et al.* Quantum dots versus organic dyes as fluorescent labels. *Nature Methods* **5**, 763–775 (2008).
24. Waring, D. R. & Hallas, G. (eds) *The Chemistry and Application of Dyes* Ch. 2 (Plenum, 1990).
25. Oliver, J., Lee, W., Park, S. & Lee, H. N. Improving resolution of miniature spectrometers by exploiting sparse nature of signals. *Opt. Express* **20**, 2613–2625 (2012).
26. Kurokawa, U., Choi, B. I. & Chang, C. C. Filter-based miniature spectrometers: spectrum reconstruction using adaptive regularization. *IEEE Sens. J.* **11**, 1556–1563 (2011).
27. Vigneau, E., Devaux, M. F., Qannari, E. M. & Robert, P. Principal component regression, ridge regression and ridge principal component regression in spectroscopy calibration. *J. Chemometr.* **11**, 239–249 (1997).
28. Bangalore, A. S., Shaffer, R. E. & Small, G. W. Genetic algorithm-based method for selecting wavelengths and model size for use with partial least-squares regression: application to near-infrared spectroscopy. *Anal. Chem.* **68**, 4200–4212 (1996).
29. Chuang, C.-H. M., Brown, P. R., Bulović, V. & Bawendi, M. G. Improved performance and stability in quantum dot solar cells through band alignment engineering. *Nature Mater.* **13**, 796–801 (2014).
30. Chin, C. D., Linder, V. & Sia, S. K. Lab-on-a-chip devices for global health: past studies and future opportunities. *Lab Chip* **7**, 41–57 (2007).

**Acknowledgements** The inception, experiments and initial analysis in this work was funded by the ARO through the Institute for Soldier Nanotechnologies (W911NF-07-D-0004). During further analysis and modelling, J.B. was supported by Tsinghua University and the Division of Physics, Mathematics and Astronomy at the California Institute of Technology.

**Author Contributions** J.B. designed the experiments with contributions from M.G.B. J.B. performed the experiments. Both authors discussed the results. J.B. wrote the manuscript with contributions from M.G.B.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.B. ([bao@tsinghua.edu.cn](mailto:bao@tsinghua.edu.cn)).



## METHODS

**Additional experimental information.** The 195 different CQD materials were taken from approximately 24 CdS and CdSe preparations using procedures from the literature<sup>31–34</sup> and by taking aliquots during CQD growth. These CQD materials were purified according to the procedures reported with their syntheses. The fluorescent emission of the CQDs was quenched by the addition of a small amount of *p*-phenylenediamine<sup>35</sup> to prevent artefacts produced by emitted light from the CQDs. To convert them into CQD filters, the CQDs were concentrated and mixed with a polyvinyl butyral (PVB) chloroform solution<sup>36</sup>. By placing a drop of the CQD/PVB solution onto a substrate and allowing evaporation of the chloroform, a CQD filter consisting of CQDs dispersed in PVB polymer (as shown in Fig. 3a) is produced. Transmission spectra for all 195 CQD filters (Fig. 3a, Extended Data Fig. 1) were measured by using a broadband light source (Deuterium Tungsten Halogen DH-2000 light source from Ocean Optics) as a reference and a commercial spectrometer (QE65000 spectrometer from Ocean Optics, approximately 0.8 nm spectral bins).

CQD filter arrays were fabricated by tightly printing small drops of CQD/PVB solutions close together onto a glass cover slip with an automatic pipette. To make the quantum dot spectrometer shown in Fig. 3c, CQD/PVB solutions were printed into an array whose dimensions were designed to match the size of a CCD; the size of each filter is about 0.5 mm across. The CCD is a SONY ICX274AL sensor with a chip size of 8.5 mm × 6.8 mm (horizontal times vertical), which was configured to provide 8-bit image data by Sensor Technologies America. The quantum dot spectrometer shown in Fig. 3c was formed by coupling the CQD filter array onto the CCD, with each CQD filter covering multiple CCD pixels. The transmission spectra of these CQD filters were characterized by comparing monochromatic light intensities recorded by the CCD with and without the CQD filters. The monochromatic light source was created by passing the light from an Oriol Research Xenon Lamp from Newport Corporation through a Triax 320 monochromator from Horiba. The output was then directed into an integrating sphere. The CCD sensor array was placed at the exit of the integrating sphere. The monochromatic light had a spectral bandwidth of 2 nm and was tuned to cover a spectral range from 390 nm to 690 nm in 1 nm steps. The transmission spectra of all individual CQD filters were measured together, one wavelength at a time. Specifically, before coupling the CQD filter array to the CCD, a reference was obtained by scanning the monochromatic light (as described above) and measuring the intensities at each pixel of the CCD. The same procedures were repeated after coupling the filter array to the CCD, resulting in a pair of intensities for each pixel for each wavelength. The ratio of each pair of intensities represents the transmission of one CQD filter at that wavelength. The transmission spectra of the set of CQD filters can be obtained from the set of ratios. Example transmission spectra of CQD filters in the array are shown in Extended Data Fig. 2.

Reference spectra in Fig. 4a–d, g, h were measured by an HR2000 spectrometer from Ocean Optics. Reference spectra of CQD fluorescent emission in Fig. 4f were measured by a Horiba Fluoromax-3 spectrofluorometer.

The standard measurement error  $\sigma$  is 0.022, which corresponds to the standard deviation of the 195  $I_i$  values obtained from the measurements subtracted from the 195  $I_i$  values that were computationally derived from equation (1) on the basis of reference spectra.

*p*-Phenylenediamine is sensitive to certain oxidation agents, which could result in filter transmission changes<sup>37</sup>. Encapsulation by polymers such as polyvinyl butyral, polymethyl methacrylate, polystyrene and polylactic acid, may offer a layer of protection. Further improvements may be made by replacing *p*-phenylenediamine with agents that are more resistant to oxidation and using CQDs with intrinsically low luminescent quantum yield such as those with a large amount of trap states<sup>19</sup>.

**Characterization of spectral resolving capability.** To characterize the spectral resolving capability of the quantum dot spectrometer, the spectrometer was tested by getting it to attempt to resolve closely positioned doublet peaks, with various peak separations (from 2 nm to 5 nm). These doublet peaks were also created using the Xenon Arc Lamp and the Triax 320 monochromator by producing a double exposure, with the monochromator tuned between two wavelengths and a 2 nm bandwidth for each monochromatic light generated by this system. The doublet peaks were characterized by the HR2000 Ocean Optics spectrometer as a reference (Extended Data Fig. 3, left panels). In this case, the quantum dot spectrometer was able to resolve two peaks at a separation of 2–3 nm (Extended Data Fig. 3, right panels).

**Effects of measurement errors on spectral resolving capability.** In the ideal case, when there is no measurement error, a quantum dot spectrometer resolves an unknown spectrum by solving a set of linear equations with a unique solution. The spectral range of the spectrometer depends on the wavelength coverage of the set of CQD filters used. The spectral resolution is determined by the number of different CQD filters and limited by the resolution of the calibrating system.

In practice, measurement errors and instrument noise is inevitable, and only approximate solutions are possible. In such a case, the spectral range is in principle not affected, but the spectral resolution decreases as the level of measurement error increases. The following simulation demonstrates this relationship.

For this simulation, a spectral range of 390–620 nm was selected. The transmission spectra of the CQD filters  $T_i(\lambda)$  are the same spectra as shown in Extended Data Fig. 1, and were binned into approximately 1.6 nm intervals (147 data points over the 230 nm range). Given these bins and the requirement of at least three points to represent a single peak, the smallest spectral distance between two distinguishable simulated peaks is 3.2 nm.

The simulations were performed according to the following procedure. An arbitrary light spectrum  $\Phi(\lambda)$  was generated. The spectrum consists of three resolution-limited peaks (Extended Data Fig. 4a), which are spectrally distinguishable from one another. Two of these peaks are separated by 3.2 nm, the smallest possible spectral distance before the two peaks merge into one broader peak. The intensities  $I_i$  of this light spectrum after passing through each CQD filter were simulated according to the following equation:

$$I_i = (1 + \epsilon_{\text{rand},i}) \sum_{\lambda} \Phi(\lambda) T_i(\lambda) \quad (2)$$

where  $\epsilon_{\text{rand},i}$  are random numbers, one for each filter, sampled from a normal distribution centred at zero, with  $\sigma = 0, 0.0001, 0.001, 0.01, 0.1$  representing different (measurement) error levels. 147 out of the 195 CQD spectra are used to generate 147  $I_i$  in the following simulations. These  $I_i$  are substituted into equation (1) to reconstruct the simulated light spectrum using least-squares linear regression. For each error level, 100 individual simulations were performed and the reconstructed spectra were averaged.

At the error level of  $\sigma = 0$ , the simulation yields a reconstructed spectrum that matches the original light spectrum perfectly (Extended Data Fig. 4b) and the two peaks separated by 3.2 nm are resolved. At the error level of  $\sigma = 0.0001$ , the simulation shows that the difference between the reconstructed spectrum and the original is extremely small (Extended Data Fig. 4c) and the two peaks separated by 3.2 nm are still resolved. As the error level increases to  $\sigma = 0.001$ , the differences between the reconstructed and the original spectrum increase (Extended Data Fig. 4d). The positions of the three peaks are nevertheless accurately represented. When the error level reaches  $\sigma = 0.01$ , the simulation no longer resolves the two peaks that are separated by 3.2 nm, but rather treats them as one broader peak (Extended Data Fig. 4e). As the error level increases to  $\sigma = 0.1$ , the error becomes large enough to lose spectral accuracy (Extended Data Fig. 4f). Nevertheless, major peak information can still be obtained from the simulation. As the peaks become broader in the reconstructed spectrum, the intensities drop, keeping the total intensity (summed over the spectral range) consistent within the tolerance of the level of the error.

To demonstrate the distribution of the 100 simulations that were averaged to produce the reconstructed spectra in Extended Data Fig. 4 (red dashed lines), we show the distribution of the 100 simulations for Extended Data Fig. 4d in Extended Data Fig. 5b. Specifically, we define  $P = (P_1 + P_2)/2$ , where the  $P_1$  and  $P_2$  are intensity values in each of the 100 simulations averaged to produce the red spectrum, at the position of the two left-most peaks in the green spectrum; we define  $V$  as the corresponding intensity value between the two peak positions and  $R = V/P$ . If  $R < 1$ , then the simulation produces two distinguishable peaks at  $P_1$  and  $P_2$ , whereas if  $R \geq 1$ , then  $P_1$  and  $P_2$  become indistinguishable in the simulation. We calculated  $R$  for each of the 100 simulations, and plotted it in a histogram (Extended Data Fig. 5b). The frequency decays rapidly with increasing  $R$ , as seen from the histogram. Of the 100 simulations, there were 60 with  $R < 1$  and 40 with  $R \geq 1$ .

To explore the resolving capability of the system at error levels of  $\sigma = 0.001, 0.01, 0.1$ , we generated more spectra (Extended Data Fig. 6a, c, e) that are similar to the spectrum shown in Extended Data Fig. 4a, but with slightly broader peaks and larger peak separations (about 4.8 nm, 5.6 nm and 13 nm, respectively) between the two left-most peaks. When the error level is at  $\sigma = 0.001$ , the simulation produces very accurate reconstructed spectra (Extended Data Fig. 6b) and the two peaks separated by about 4.8 nm were easily resolved. At the error level of  $\sigma = 0.01$ , the simulation produces a reconstructed spectrum in which the two peaks separated by about 5.6 nm separation are just resolved (Extended Data Fig. 6d). At the error level of  $\sigma = 0.1$ , the simulation suggests that the spectral resolving capability is notably reduced and only two peaks separated by more than about 13 nm are resolved (Extended Data Fig. 6f).

**Effects of the number of filters on spectral resolving capability.** A remedy to the loss of spectral information and the reduced spectral resolving capability under conditions with relatively high levels of measurement error and instrument noise is to increase the number of CQD filters and thus the number of intensities  $I_i$ . This is because when redundant spectral information is provided by the extra filters, the

error level is statistically reduced. In addition, the algorithm may be modified to reject data points with errors larger than some defined level.

In the following simulations, the number of CQD filters is increased from 147 to 195. The same procedures and original (incident) spectra are used as in the above simulations. The spectra reconstructed with 195 filters demonstrated improvements from those using 147 filters (Extended Data Fig. 7). However, the limited number of CQD filters produced in the experiments limits our ability to run further simulations with increased filter numbers, which we expect would demonstrate further improvements in resolution.

**Effects of algorithms on spectral reconstruction accuracy.** The example shown below demonstrates the difference that a more sophisticated algorithm can make to the reconstruction of a spectrum. The red crosses in Extended Data Fig. 8a correspond to the reconstructed spectrum using the measurements from the quantum dot spectrometer; the blue line corresponds to the reference spectrum taken by the Ocean Optics spectrometer. In this case, the spectral reconstruction was performed using a least-squares linear regression. However, with a more sophisticated algorithm, based on generalized least squares and regularization, the same data are reconstructed to give the spectrum that is plotted with the red markers in Extended Data Fig. 8b (the blue line is the same reference spectrum). This shows that such algorithms reduce ghost (spurious) lines and therefore increase the reconstruction accuracy.

**Effects of measurement errors on dynamic range.** The dynamic range of a spectrometer is defined as the ratio between the maximum intensity it can measure and the minimum signal that is distinguishable from the noise. In a grating-based spectrometer, the maximum intensity is typically equal to the maximum range of the analogue-to-digital converter (ADC) of the detector. The minimum signal that is distinguishable from the noise is represented by the standard deviation of the dark signal (that is, the dark counts of the detector).

In a quantum dot spectrometer, however, the dynamic range should be calculated on the basis of the maximum intensity and the dark counts from the reconstructed final spectrum, instead of those read directly from the detector. Therefore, the dynamic range is not only determined by the range of the ADC and the dark signal, but also by the reconstruction algorithm and the measurement errors. In the following simulation, we show how the level of the measurement errors affects the dynamic range of a quantum dot spectrometer.

We consider a case in which a photodetector with a 16-bit ADC (which can produce an integer from 0 to  $2^{16} - 1 = 65,535$ ) is used. Similar to the above simulations, a spectral range of 390–620 nm is selected, the transmission spectra  $T_i(\lambda)$  of the 195 CQD filters are used and binned to have approximately 1.6 nm intervals (147 data points over the 230 nm range). For the maximum intensity, we consider an arbitrary incident-light spectrum  $\Phi(\lambda)$  of a solitary point peak with an intensity of 65,535. The intensities  $I_i$  of  $\Phi(\lambda)$  after passing through each CQD filter were computed according to equation (2). These  $I_i$  are substituted into equation (1) to reconstruct the simulated light spectrum using least-squares linear regression.  $I_{\max}$  is the maximum intensity in the reconstructed spectrum. This procedure is repeated 100 times at each error level. The 100  $I_{\max}$  are averaged to give  $I_{\max, \text{mean}}$ .

To calculate the standard deviation of the dark signal, we generate a series of dark counts  $I_{d,i} = \lfloor 65,535(\sigma + e_{\text{rand},d,i}) \rfloor$ , where  $e_{\text{rand},d,i}$  are random numbers, one for each filter, sampled from a normal distribution, assuming both a standard deviation of  $\sigma$  and a mean of  $\sigma$ , with  $\sigma = 0.0001, 0.001, 0.01, 0.1$  representing different error levels. The absolute value is taken to avoid any negative detector

counts in this simulation. After generating the dark counts, the spectrum is reconstructed on the basis of these  $I_{d,i}$ .  $\sigma_{\Phi,d}$  is the standard deviation of the reconstructed spectral intensities across the entire wavelength range. This procedure is repeated 100 times at each error level, and the 100  $\sigma_{\Phi,d}$  are averaged to give  $\sigma_{\Phi,d, \text{mean}}$ .

The dynamic range at each error level is  $\text{DR} = I_{\max, \text{mean}} / \sigma_{\Phi,d, \text{mean}}$ . Given in Extended Data Table 1 are the simulated  $I_{\max, \text{mean}}$ ,  $\sigma_{\Phi,d, \text{mean}}$  and DR for a quantum dot spectrometer for different levels of measurement errors. This shows that the dynamic range of a quantum dot spectrometer increases as the level of measurement errors decreases.

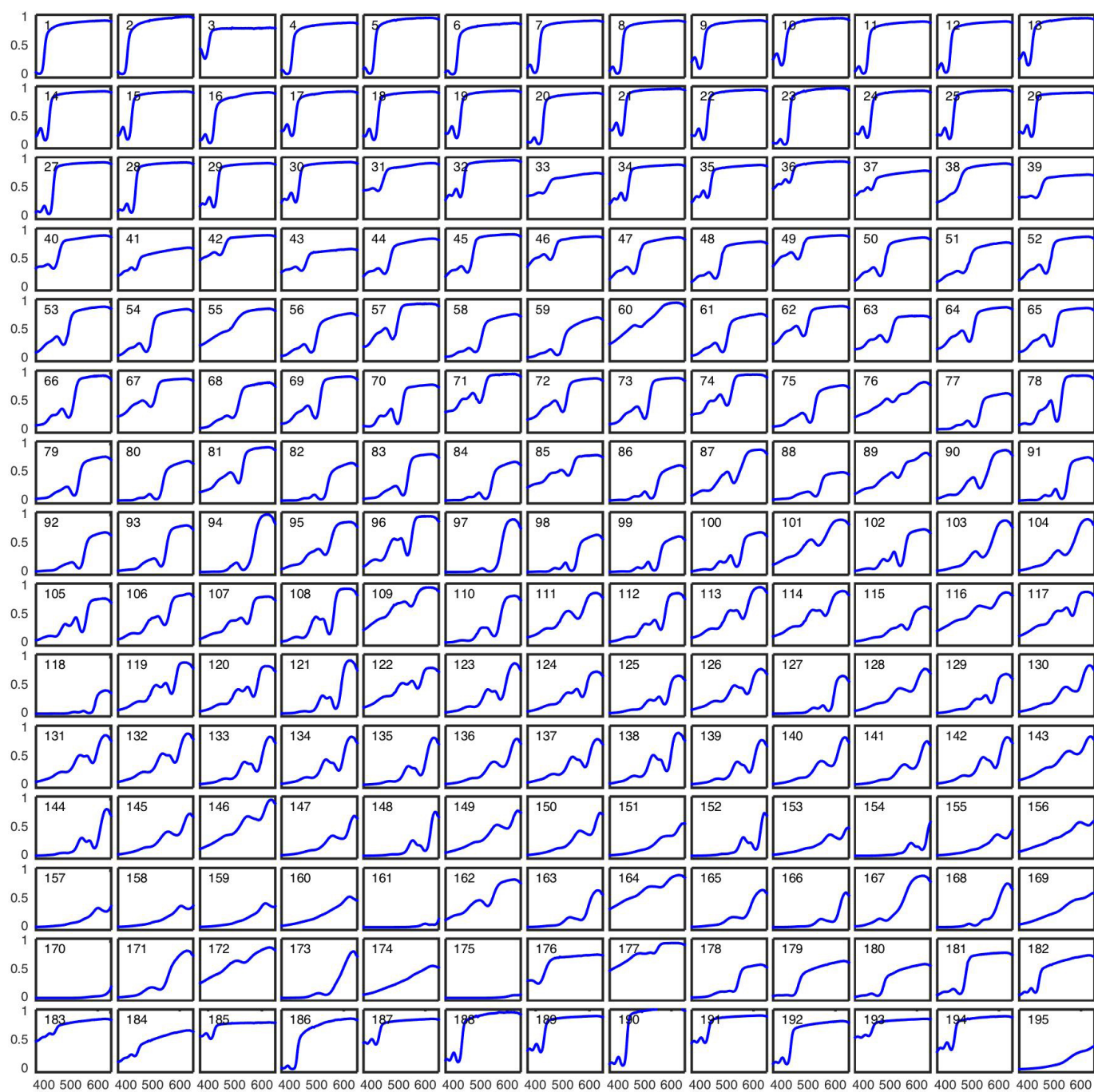
When the system is Poisson-noise limited, the dynamic range may be different to the above simulations because Poisson noise is sampled from a Poisson distribution instead of a normal distribution. However, the normal distribution is a good approximation of the Poisson distribution when the mean (photon number in this case) is large enough ( $> 1,000$ , for instance). Typically, when looking at the dynamic range, the signal level (number of photons) is relatively high. Therefore, we expect that the dynamic range of a Poisson-noise dominated quantum dot spectrometer follows the same trend as for the above simulations.

For example, consider a system with gain of one electron per analogue-to-digital unit (ADU) and a quantum efficiency of 1/3. In this system, a 65,535 ADU would imply 196,605 photons, which corresponds to a standard error of about 443. Assuming the errors are Poisson distributed, we calculated  $I_{\max, \text{mean}} \approx 64,500$  using the same methods as described above. This result is indistinguishable from the result for a normal distribution at an error level of  $\sigma = 0.0022 \approx 443/196,605$ .

**Stability analysis.** The stability of the quantum dot spectrometer was tested under typical ambient laboratory conditions. Plotted in Extended Data Fig. 9 are two spectra measurements taken about six months apart, without recalibration. The peaks shown in both plots are at 400 nm, 450 nm, 500 nm, 501 nm, 503 nm, 505 nm, 550 nm and 600 nm; both sets of results reproduce the peaks accurately.

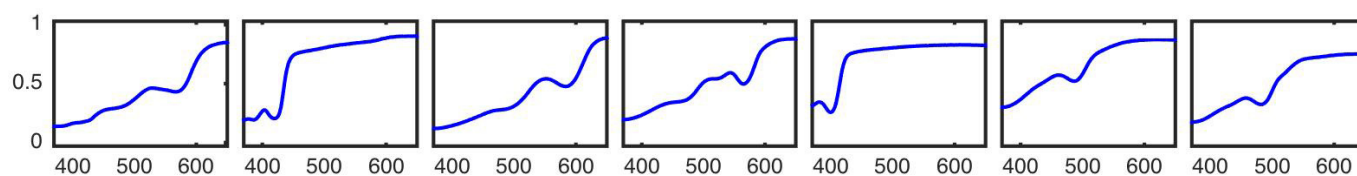
In addition, it is known that absorption is one of the most stable properties of quantum dots. Commercial products based on emission of quantum dots<sup>38</sup> and the recent work on air-stable quantum dot solar cells<sup>29</sup> showed that high stability of the more sensitive properties such as fluorescence emission and charge transportation can nevertheless be achieved by carefully designing the chemistry of quantum dots and the conjugating materials.

31. Chen, O. *et al.* Synthesis of metal-selenide nanocrystals using selenium dioxide as the selenium precursor. *Angew. Chem. Int. Ed.* **47**, 8638–8641 (2008).
32. Qu, L., Peng, Z. A. & Peng, X. Alternative routes toward high quality CdSe nanocrystals. *Nano Lett.* **1**, 333–337 (2001).
33. Peng, Z. A. & Peng, X. Formation of high-quality CdTe, CdSe, and CdS nanocrystals using CdO as precursor. *J. Am. Chem. Soc.* **123**, 183–184 (2001).
34. Cao, Y. C. & Wang, J. One-pot synthesis of high-quality zinc-blende CdS nanocrystals. *J. Am. Chem. Soc.* **126**, 14336–14337 (2004).
35. Sharma, S. N., Pillai, Z. S. & Kamat, P. V. Photoinduced charge transfer between CdSe quantum dots and *p*-phenylenediamine. *J. Phys. Chem. B* **107**, 10088–10093 (2003).
36. Norris, D. J., Sacra, A., Murray, C. B. & Bawendi, M. G. Measurement of the size dependent hole spectrum in CdSe quantum dots. *Phys. Rev. Lett.* **72**, 2612–2615 (1994).
37. Ćirić-Marjanović, G. *et al.* The oxidative polymerization of *p*-phenylenediamine with silver nitrate: toward highly conducting micro/nanostructured silver/conjugated polymer composites. *J. Polym. Sci. A* **49**, 3387–3403 (2011).
38. Shirasaki, Y., Supran, G. J., Bawendi, M. G. & Bulović, V. Emergence of colloidal quantum-dot light-emitting technologies. *Nature Photon.* **7**, 13–23 (2013).

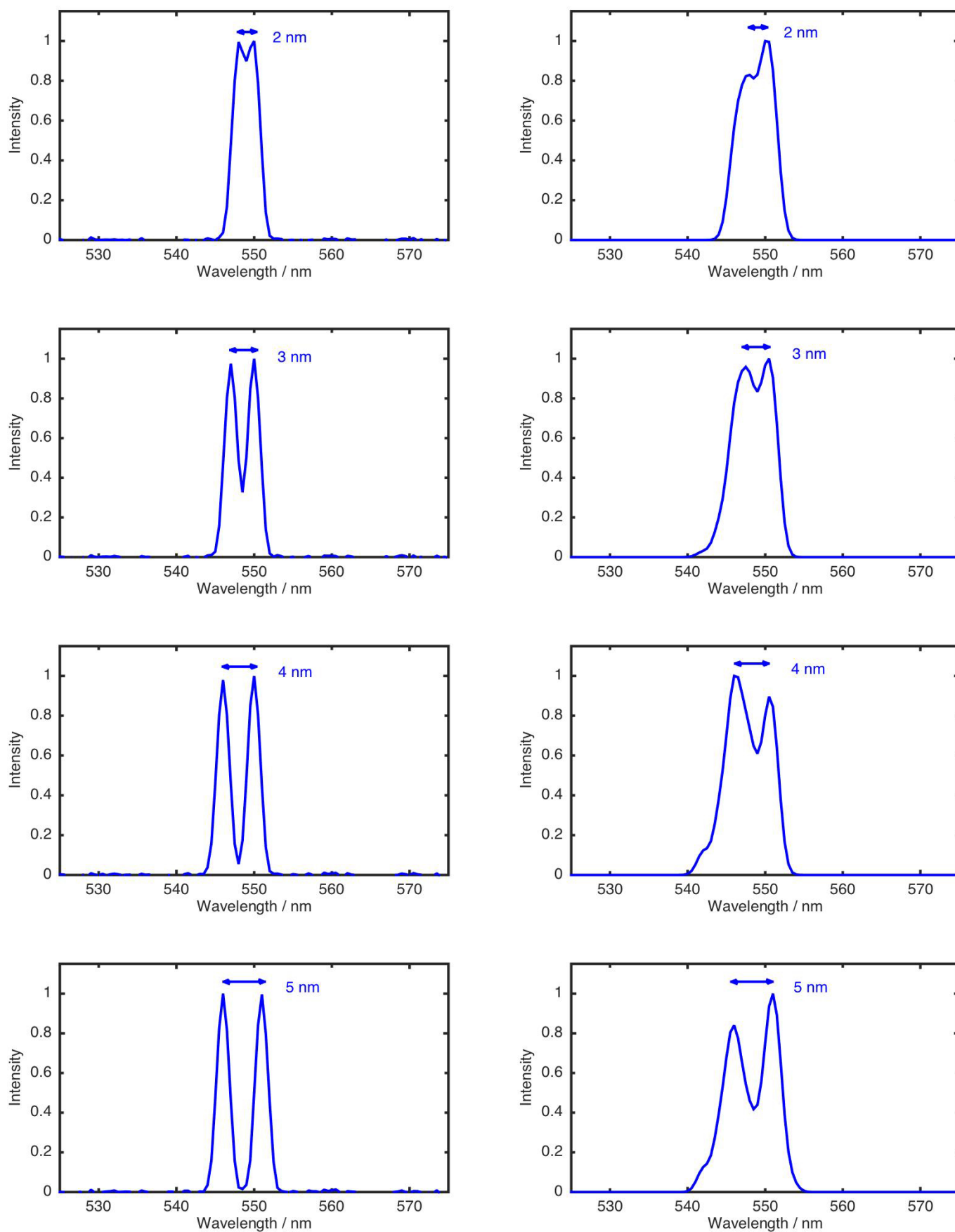


**Extended Data Figure 1 | Transmission spectra for all 195 CQD embedded thin films shown in Fig. 3a.** In each subplot, the horizontal axis measures wavelength (in nanometres) and the vertical axis measures the transmission fraction ( $\times 100\%$ ).



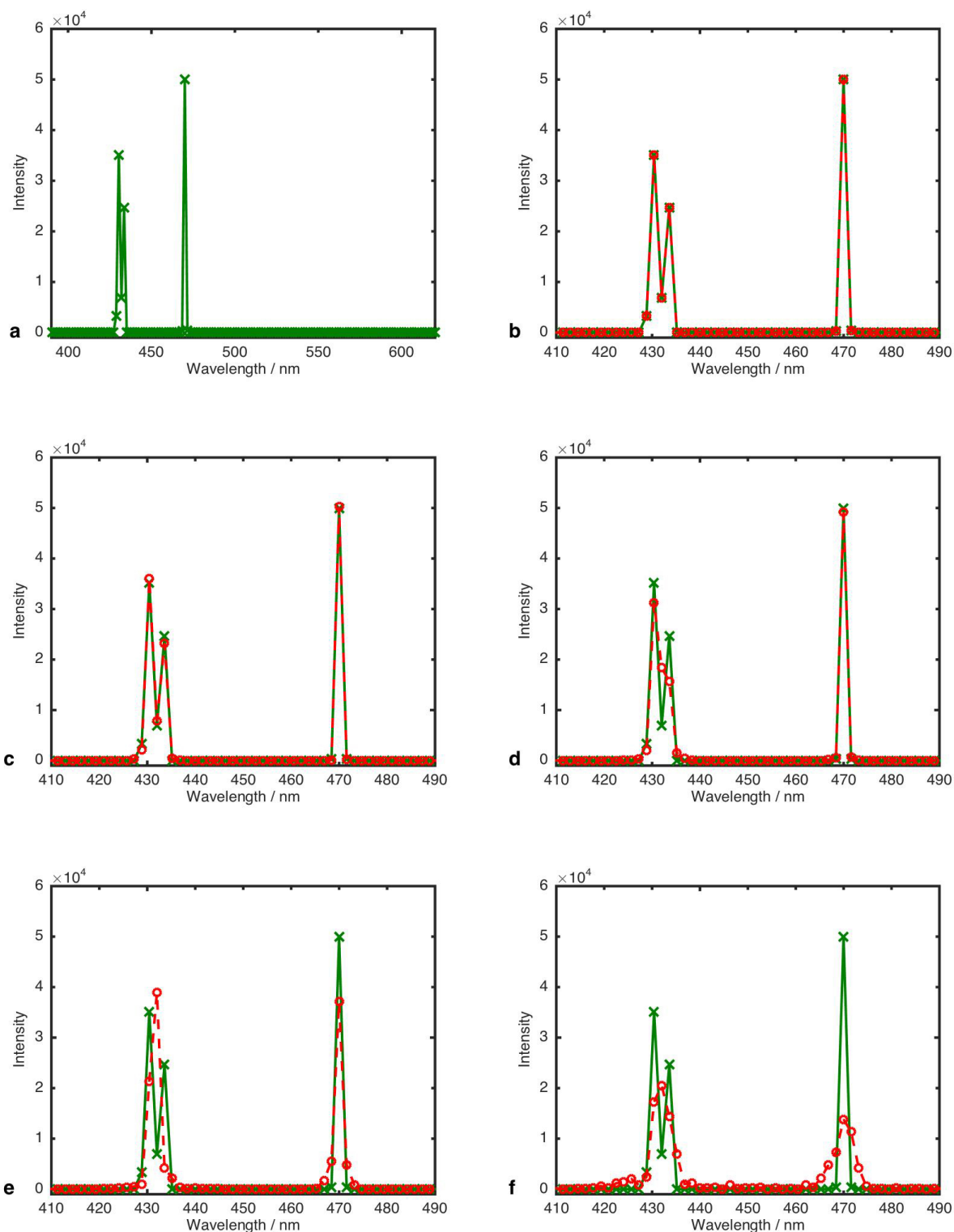


**Extended Data Figure 2 | Transmission spectra for selected CQD filters in the CQD filter array.** In each subplot, the horizontal axis measures wavelength (in nanometres) and the vertical axis measures the transmission fraction ( $\times 100\%$ ).



**Extended Data Figure 3 | Characterization of the spectral resolution of the quantum dot spectrometer.** The left panels are doublet peaks with peak separations of 2–5 nm measured by a HR2000 spectrometer as a reference. The

right panels are the same doublet peaks measured by the quantum dot spectrometer.

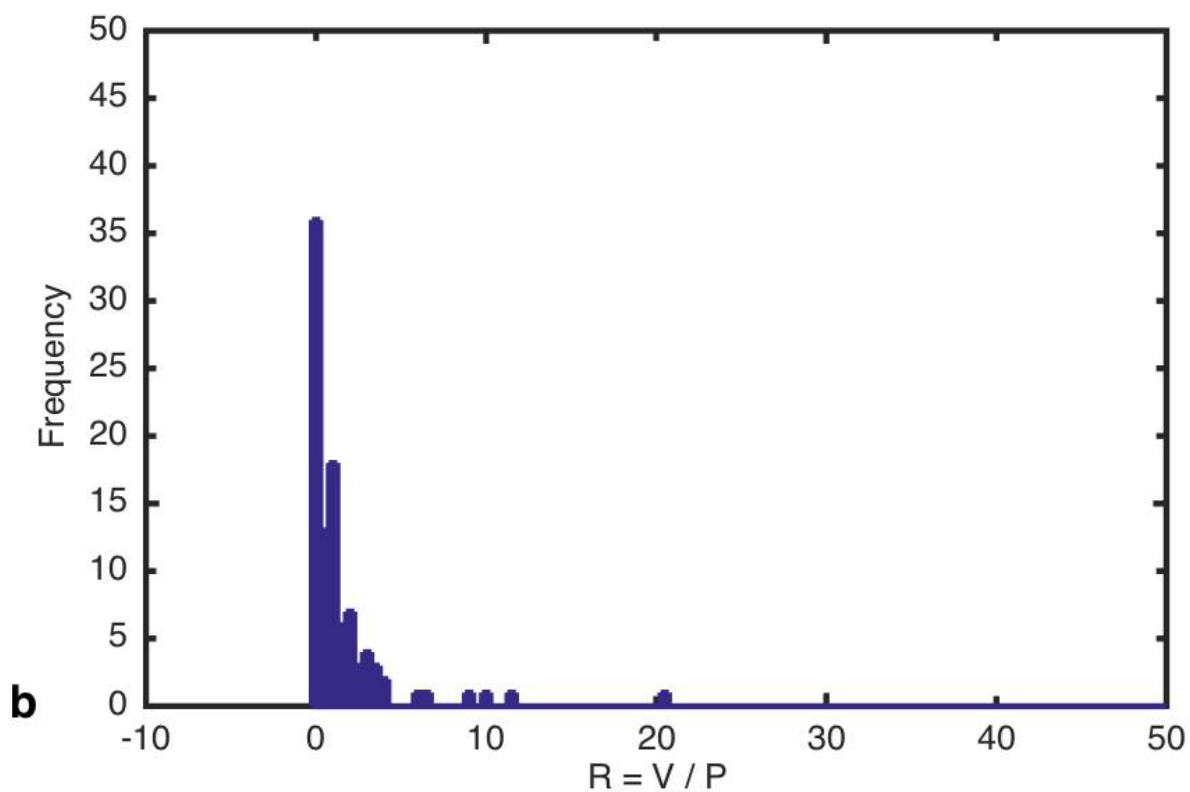
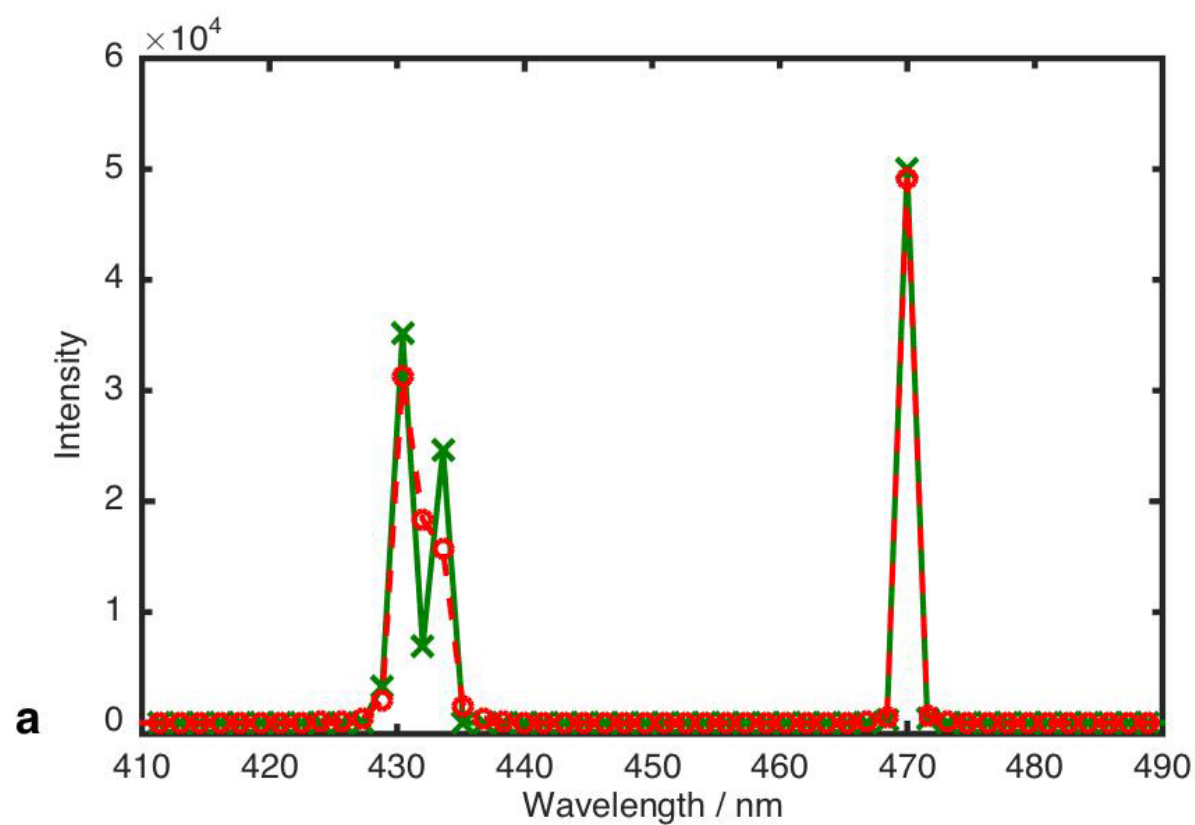


#### Extended Data Figure 4 | A simulated spectrum and spectral reconstruction.

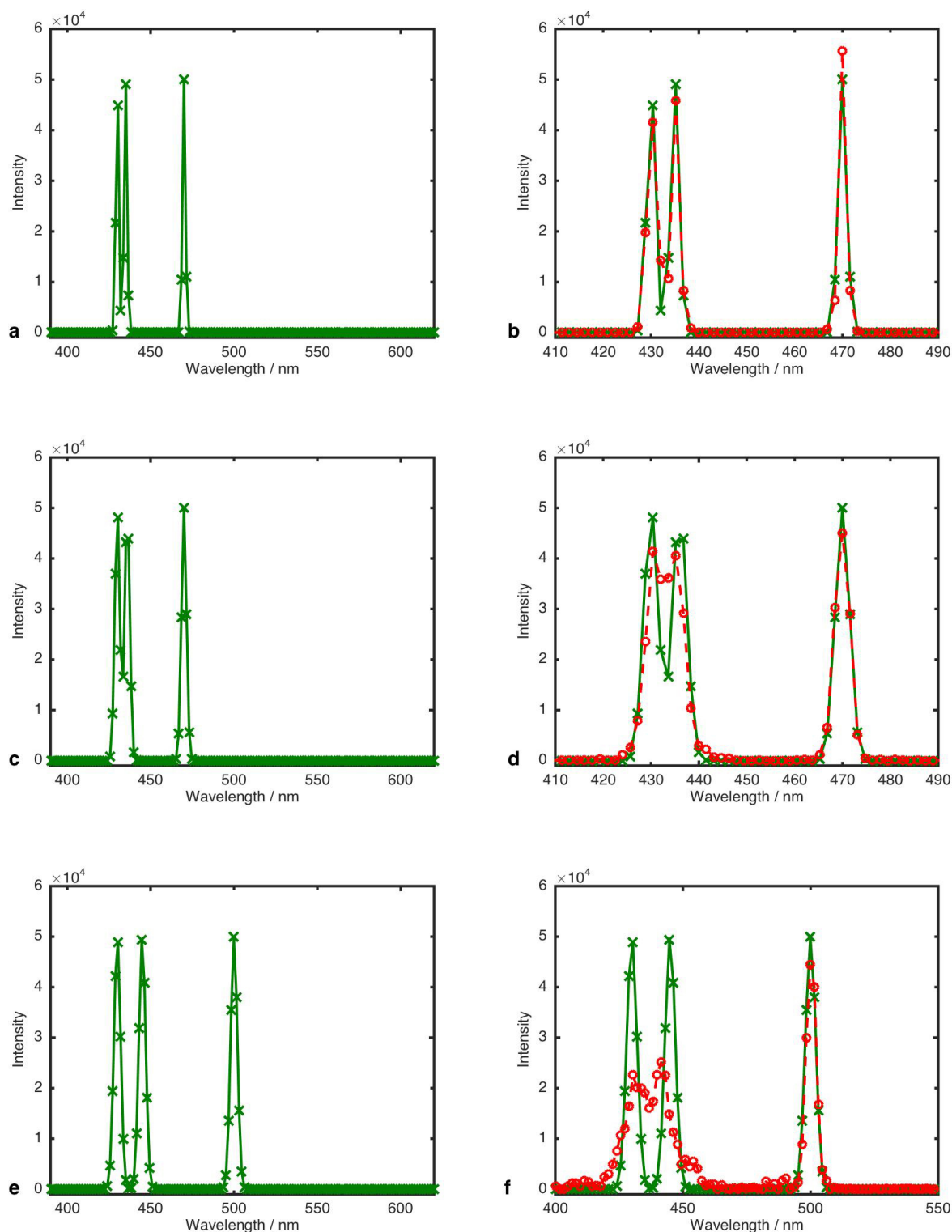
**a**, The green crosses represent the data points of a simulated light spectrum in intervals of 1.6 nm. The separation between the two left-most peaks is 3.2 nm. **b**, Spectral reconstruction ( $\sigma = 0$ ,  $n_F = 147$ ). With  $\sigma = 0$ , the spectral reconstruction process is equivalent to solving a set of linear equations that has a unique solution. The reconstructed spectrum (the red dotted line with red circles representing data points) matches the original (incident) light spectrum perfectly and the two peaks separated by 3.2 nm are resolved. **c–f**, Spectral reconstruction ( $\sigma = 0.0001, 0.001, 0.01, 0.1$ , respectively,

$n_F = 147$ ) using least-squares linear regression. **c**, The original light spectrum is reproduced accurately and the two peaks separated by 3.2 nm are resolved. **d**, The original light spectrum is reproduced reasonably well and the peak positions accurately represented. **e**, The spectral reconstruction does not reproduce the original light spectrum very accurately and some spectral information is lost. **f**, The spectral reconstruction no longer reproduces the original light spectrum accurately and a lot of spectral information is lost. Nevertheless, major peak information can still be obtained from the simulation.



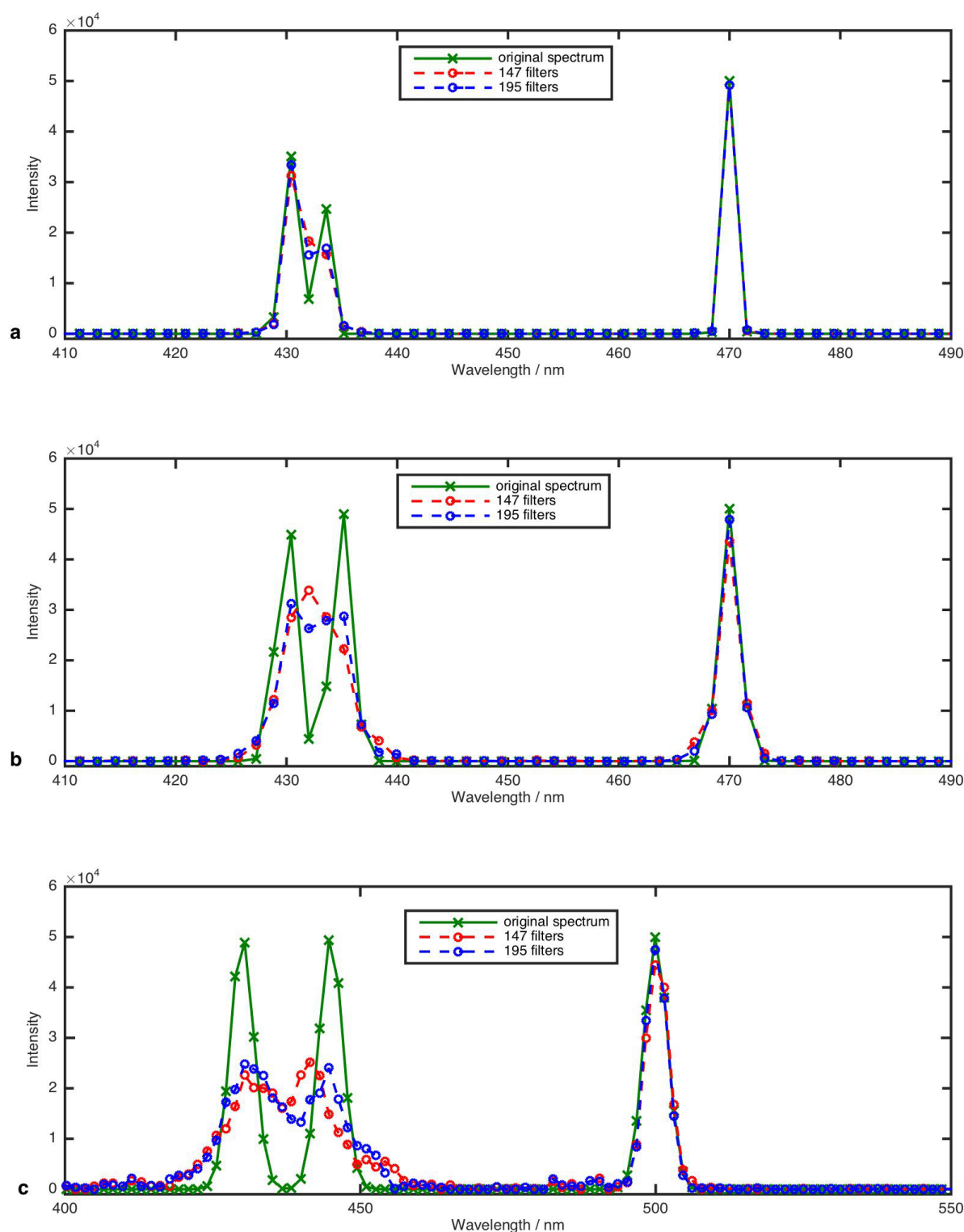


**Extended Data Figure 5 | Distribution of simulations.** **a**, Same as Extended Data Fig. 4d. **b**, The frequency (of the given value of  $R$  occurring in the 100 simulations averaged to produce **a**) decay with increasing  $R$ .



**Extended Data Figure 6 | Simulated spectra and spectral reconstructions.**  
**a, c, e**, Second, third and fourth simulated light spectra. The green crosses represent the simulated data points, in intervals of 1.6 nm. The separation between the two leftmost peaks is about 4.8 nm, 5.6 nm and 13 nm, respectively.  
**b, d, f**, Spectral reconstruction ( $\sigma = 0.001, 0.01, 0.1$ , respectively,  $n_F = 147$ ) using least-squares linear regression. **b**, The original light spectrum is

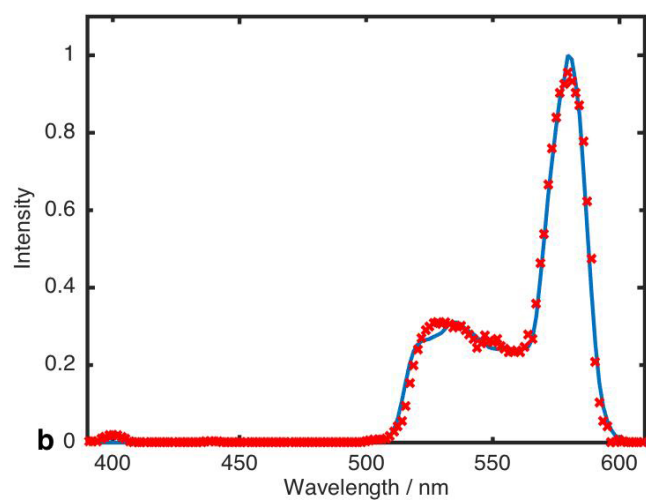
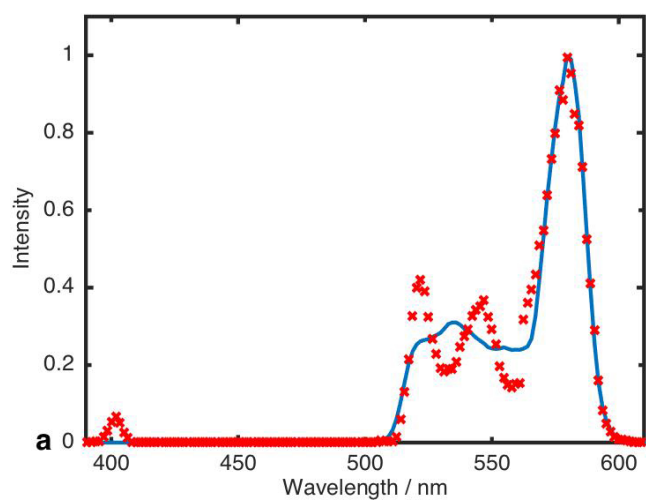
reproduced very accurately and little spectral information is lost. The two leftmost peaks are resolved. **d**, Spectral reconstruction does not reproduce the original light spectrum very accurately and some spectral information is lost. However, the two leftmost peaks are resolved. **f**, Spectral reconstruction no longer reproduces the original light spectrum very accurately and some spectral information is lost. However, the two leftmost peaks are resolved.



**Extended Data Figure 7 | Comparison of spectral reconstructions with  $n_F = 147$  and  $n_F = 195$ .** a–c, The separation between the two left-most peaks is 3.2 nm, 4.8 nm and 13 nm, respectively. The spectral reconstruction ( $\sigma = 0.001, 0.01, 0.1$ , respectively) using both filter sets reproduced the main

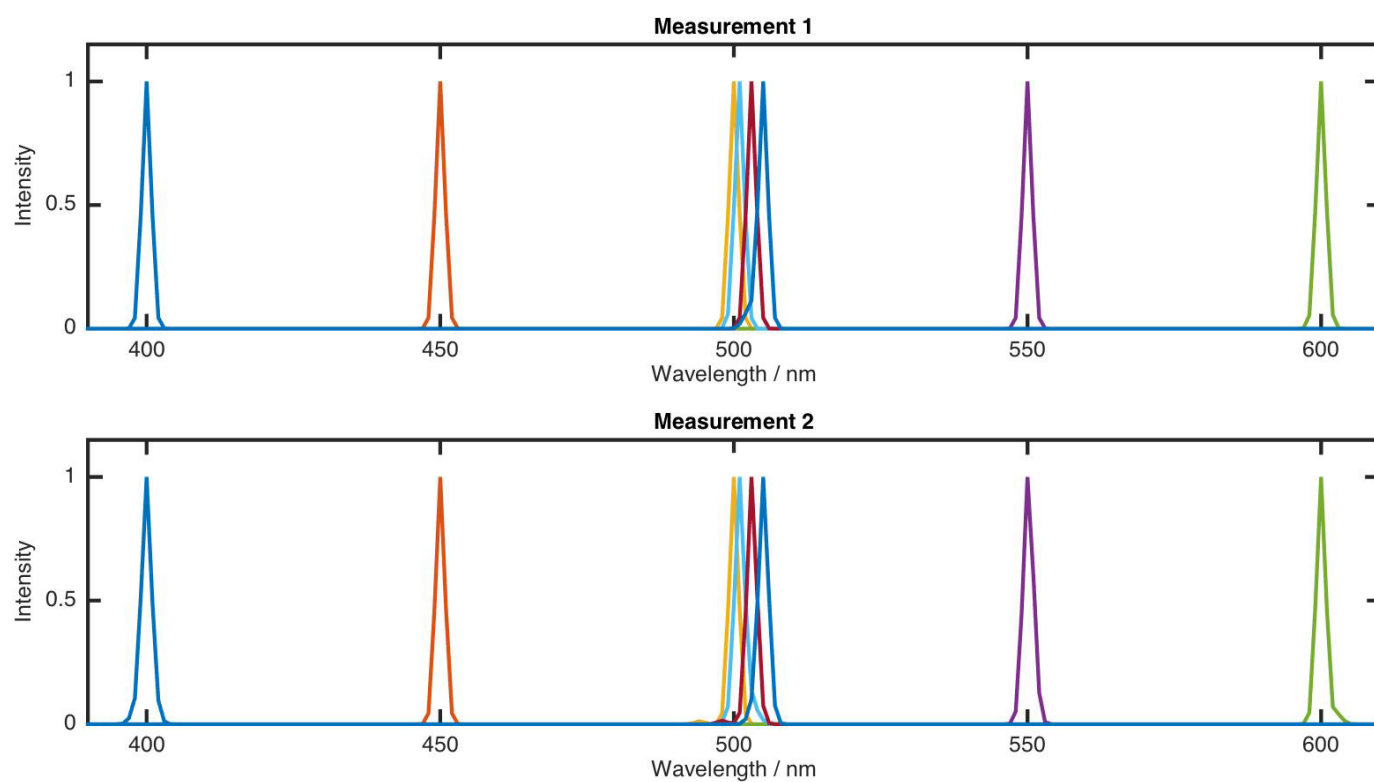
features of the original light spectrum. An improvement in the reconstructed spectrum with 195 filters is observed, compared to that with 147 filters, where all other conditions are the same. However, due to the small relative difference between 147 and 195, the improvement is limited.





**Extended Data Figure 8 | Effects of algorithms on spectral reconstruction accuracy.** **a**, Plotted with the red markers is a spectrum measured by the quantum dot spectrometer and reconstructed using least-squares linear regression. The blue line represents the reference spectrum. **b**, Plotted with the

red markers is the spectrum based on the same quantum dot spectrometer measurement data, but reconstructed using a more sophisticated algorithm. The blue line represents the same reference spectrum.



**Extended Data Figure 9 | Stability analysis.** In both Measurement 1 and Measurement 2 (taken six months apart, without recalibration), the peaks shown are at 400 nm, 450 nm, 500 nm, 501 nm, 503 nm, 505 nm, 550 nm and 600 nm.

Extended Data Table 1 | Simulation of the effects of errors on the dynamic range

| Error Level / $\sigma =$        | 0.0001 | 0.001 | 0.01  | 0.1   |
|---------------------------------|--------|-------|-------|-------|
| $I_{\max, \text{mean}}$         | 65000  | 65000 | 62000 | 48000 |
| $\sigma_{\Phi, d, \text{mean}}$ | 2.4    | 24    | 240   | 2400  |
| DR                              | 27000  | 2700  | 260   | 20    |



# A model-tested North Atlantic Oscillation reconstruction for the past millennium

Pablo Ortega<sup>1†</sup>, Flavio Lehner<sup>2,3†</sup>, Didier Swingedouw<sup>4</sup>, Valerie Masson-Delmotte<sup>1</sup>, Christoph C. Raible<sup>2,3</sup>, Mathieu Casado<sup>1</sup> & Pascal Yiou<sup>1</sup>

The North Atlantic Oscillation (NAO) is the major source of variability in winter atmospheric circulation in the Northern Hemisphere, with large impacts on temperature, precipitation and storm tracks<sup>1</sup>, and therefore also on strategic sectors such as insurance<sup>2</sup>, renewable energy production<sup>3</sup>, crop yields<sup>4</sup> and water management<sup>5</sup>. Recent developments in dynamical methods offer promise to improve seasonal NAO predictions<sup>6</sup>, but assessing potential predictability on multi-annual timescales requires documentation of past low-frequency variability in the NAO. A recent bi-proxy NAO reconstruction<sup>7</sup> spanning the past millennium suggested that long-lasting positive NAO conditions were established during medieval times, explaining the particularly warm conditions in Europe during this period; however, these conclusions are debated. Here, we present a yearly NAO reconstruction for the past millennium, based on an initial selection of 48 annually resolved proxy records distributed around the Atlantic Ocean and built through an ensemble of multivariate regressions. We validate the approach in six past-millennium climate simulations, and show that our reconstruction outperforms the bi-proxy index. The final reconstruction shows no persistent positive NAO during the medieval period, but suggests that positive phases were dominant during the thirteenth and fourteenth centuries. The reconstruction also reveals that a positive NAO emerges two years after strong volcanic eruptions, consistent with results obtained from models and satellite observations for the Mt Pinatubo eruption in the Philippines<sup>8,9</sup>.

In the North Atlantic sector, roughly half of the interannual variability in winter atmospheric pressure is explained by fluctuations in the NAO. The NAO is characterized by a changing dipole of sea-level pressure between the Azores and Iceland, and has widespread impacts on temperature and precipitation across Eurasia, North Africa, Greenland and northern North America<sup>10</sup>.

A recent study discarded the idea that there has been a strong direct radiative influence of solar forcing on Northern Hemisphere temperatures in the past millennium<sup>11</sup>. Consequently, the hypothesis has emerged that changes in the NAO could explain part of the decadal-to-centennial variations in these temperatures. Also, NAO variability itself can be externally driven, and therefore allow for an indirect effect of external forcing on temperature not accounted for in the previous study<sup>11</sup>.

A millennial bi-proxy NAO reconstruction<sup>7</sup> (hereafter NAO<sub>Trouet</sub>) shows persistent positive phases during the Medieval Climate Anomaly (MCA, roughly AD 1000–1300; ref. 12), as does a multi-millennial NAO reconstruction from one lake sediment record in Greenland<sup>13</sup>. Other proxy records from the Iberian Peninsula<sup>14</sup> support drier conditions during this period, a common fingerprint of positive NAO phases. However, some doubts have been raised regarding the persistence of a positive NAO phase during medieval times. For example, such persistent phases are not reproduced in any of 13 different past-millennium climate simulations<sup>15</sup>. Also, if the NAO

were predominantly positive during this period, low temperatures should be recorded in southwest Greenland<sup>16</sup>, and warm temperatures in northwestern Europe, but no indication of such anomalies appears in ice-core records and documentary sources<sup>17</sup>. Proxy-based NAO reconstructions are challenging because of the nonstationarity of the statistical relationships and regional teleconnections on which they are based<sup>18,19</sup>. Moreover, a ‘perfect-model’ study, using reanalyses and climate models as physically consistent surrogates of the real world (see Methods), has shown that the two proxy records involved in NAO<sub>Trouet</sub> are insufficient to guarantee a robust reconstruction throughout the entire last millennium<sup>20</sup>.

All of these factors point towards the need for a new multi-proxy NAO reconstruction for the past millennium, based on a larger set of proxy records that is representative of several NAO fingerprints in temperature, precipitation and droughts. Unlike NAO<sub>Trouet</sub>, which is based on 30-year smoothed data, our reconstruction is built at annual resolution, allowing us to test the effect of major volcanic eruptions and the 11-year solar cycle on the NAO. Although it is well established that a positive NAO-like sea-level pressure pattern is observed one to two years after major volcanic eruptions, most analyses focus on the Pinatubo eruption<sup>8,9</sup> (the only major eruption to be observed by satellite). The effect of solar activity is less well constrained, although recent studies suggest that the Arctic Oscillation (which is closely linked to the NAO) is modulated by the 11-year solar cycle<sup>21,22</sup>.

Our study is based on 48 annually resolved proxy time series (Fig. 1a and Extended Data Table 1), which are significantly correlated ( $P$  value  $< 0.10$ ) with the longest instrumental record of the NAO (hereafter NAO<sub>Vinther</sub><sup>23</sup>) in the period in which they overlap. All of these records were screened to fulfil two additional conditions: first, they encompass the eleventh to the twentieth centuries continuously; and second, they have a documented relationship in previous literature with a climate variable easily extractable from climate models.

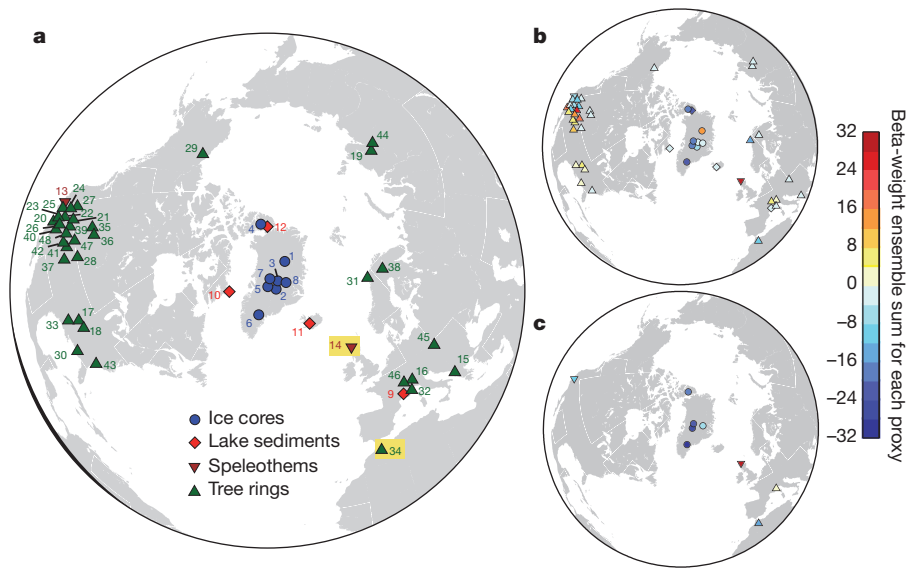
The NAO is reconstructed using principal component regression (PCR), following the procedure of a previous reconstruction from AD 1400 onwards<sup>24</sup> (see Methods). To identify the robust features in the reconstruction, we generate an ensemble that explores the sensitivity to calibration, by extracting 100 random samples of 117 years from the period that is common to all proxies and the NAO (that is, AD 1823–1969). The remaining 30 years will later be used for validation.

Two different alternative ensemble reconstructions are constructed by imposing restrictions on the initial proxy data set. The first ensemble (NAO<sub>cc</sub>, for ‘calibration-constrained’) uses as many proxy records as possible to guarantee an optimal calibration, verifying first that the proxy records still correlate significantly with NAO<sub>Vinther</sub> within the calibration period. Figure 1b illustrates the individual contribution (as cumulated weights; see Methods) of the proxy records to the NAO<sub>cc</sub> reconstruction. Each record participates at least once in the ensemble. This approach optimizes the statistical match with the

<sup>1</sup>LSCE/IPSL, UMR 8212 (CEA-CNRS-UVSQ), CEA Saclay, Gif-sur-Yvette 91191, France. <sup>2</sup>Climate and Environmental Physics, Physics Institute, University of Bern, Sidlerstrasse 5, 3012 Bern, Switzerland.

<sup>3</sup>Oeschger Centre for Climate Change Research, University of Bern, Sidlerstrasse 5, 3012 Bern, Switzerland. <sup>4</sup>EPOC, Université Bordeaux 1, Allée Geoffroy Saint-Hilaire, Pessac 33615, France.

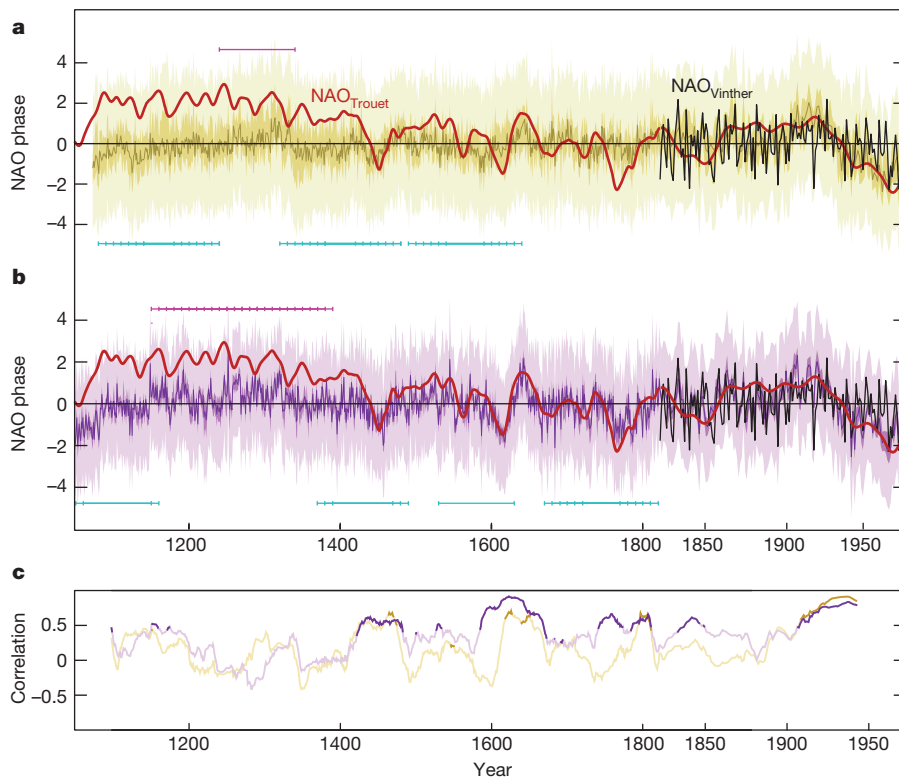
<sup>†</sup>Present addresses: NCAS Climate, Department of Meteorology, University of Reading, Reading RG6 6BB, UK (P.O.); National Center for Atmospheric Research, 1850 Table Mesa Drive, Boulder, Colorado 80305, USA (F.L.).



**Figure 1 | Proxy selection.** **a**, Location of the initial 48 proxy records (ice cores, lake sediments, speleothems or tree rings) preselected for the reconstructions, using symbols and colours to represent the different types of archive. Some symbols have been slightly displaced to improve their visibility. The two proxies used in  $NAO_{Trouet}$  are highlighted with yellow boxes. The true coordinates of each record are detailed in Extended Data Table 1. **b**, **c**, The colours of the symbols now represent the cumulated beta-weights (see Methods) associated with the proxies that contribute to the  $NAO_{cc}$  (**b**) and  $NAO_{mc}$  (**c**) ensemble reconstructions.

instrumental  $NAO_{Vinter}$  within the calibration period, but cannot ensure that  $NAO$ -proxy relationships have an actual physical basis, or that these relationships are stationary in time. Both issues are tackled simultaneously in a second ensemble reconstruction ( $NAO_{mc}$  for ‘model-constrained’). The proxy selection is first narrowed down with reanalyses (Extended Data Table 2) to guarantee realistic  $NAO$  teleconnections, and further constrained with eight past-millennium PMIP3 simulations (Paleoclimate Modelling Intercomparison Project Phase 3; Extended Data Table 3). This latter step ensures that the teleconnections are stable throughout the whole period. The  $NAO_{mc}$  ensemble reconstruction uses those proxies whose  $NAO$ -climate fingerprint is supported by at least one of the reanalyses and by half of the PMIP3 simulations. This final approach is highly restrictive and leads to a selection of only nine proxy records (Fig. 1c). If some teleconnections are unresolved by models, this method may discard proxy records that are actually sensitive to the  $NAO$ .

The two ensemble reconstructions are displayed in Fig. 2a and b and compared to  $NAO_{Trouet}$  with 50-year moving correlations (Fig. 2c). Only  $NAO_{mc}$  shows some periods of relatively good agreement with  $NAO_{Trouet}$  outside the instrumental period, coinciding with the largest-amplitude changes. However, none of the ensemble reconstructions exhibits persistent positive  $NAO$  phases in medieval times. Instead, a Student’s  $t$ -test on their means, applied in 100-year intervals, highlights a first period with non-zero mean negative values that is common to  $NAO_{mc}$  and  $NAO_{Trouet}$  (left-hand cyan horizontal bar in Fig. 2b). In AD 1150, the  $NAO_{mc}$  undergoes a rapid transition towards positive phases, which remain significantly different ( $P < 0.05$ ) from zero for about 240 years (magenta horizontal bar). In the following four centuries, until the industrial period, the  $NAO_{mc}$  again depicts predominantly negative values. Similar features, albeit with differences in the timing of phase changes, are found in the  $NAO_{cc}$  (Fig. 2a).

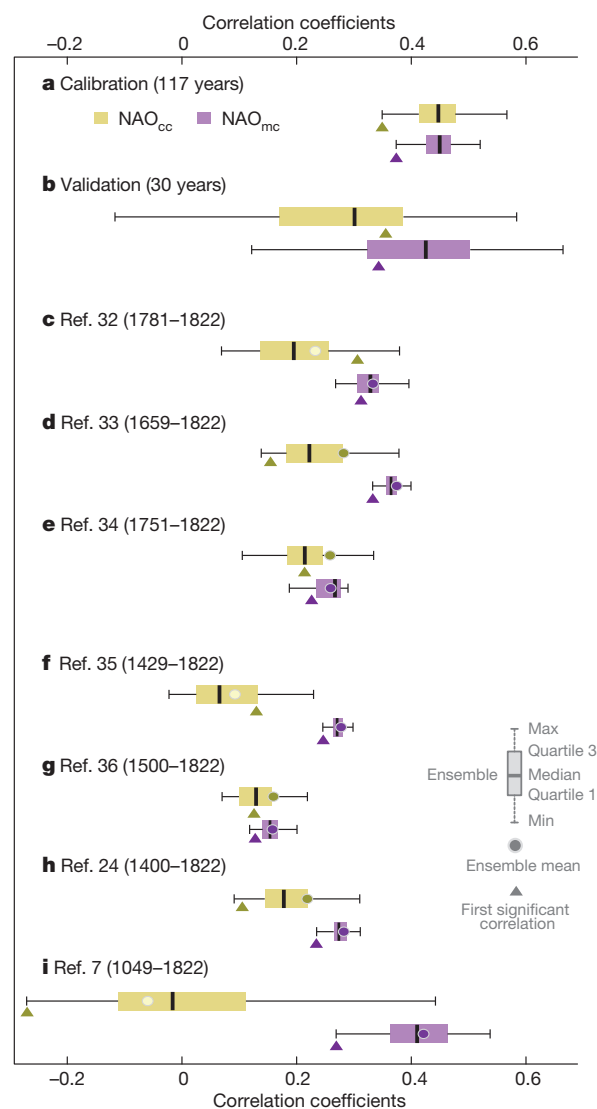


**Figure 2 | Ensemble NAO reconstructions.** **a**, Ensemble of 100 yearly  $NAO_{cc}$  reconstructions. The intermediate and light yellow envelopes represent, respectively, the total ensemble spread, and the regression uncertainties across the ensemble (as departures of  $\pm 2$  standard error, s.e., from the individual reconstructions). Also shown are the ensemble mean (dark yellow line),  $NAO_{Vinter}^{23}$  (black) and  $NAO_{Trouet}^7$  (red). Horizontal magenta (or cyan) bars enclose centuries (marked by tick marks) with positive (or negative) mean  $NAO$  values that are significantly different ( $P < 0.05$ ) than zero (see Methods). **b**,  $NAO_{mc}$  ensemble. **c**, 50-year running correlations between the mean ensemble reconstructions and  $NAO_{Trouet}$ . Dark-coloured lines represent correlations significant at the 95% confidence level ( $P < 0.05$ ).

We now assess the difference in performance between the ensembles. By construction,  $NAO_{cc}$  exhibits higher correlations with  $NAO_{vinther}$  within the calibration period than does  $NAO_{mc}$  (Fig. 3a), as  $NAO_{cc}$  includes more proxy data, and therefore more principal components contribute to the multivariate regression (Extended Data Table 4). However, this does not translate into larger correlations for the validation years (Fig. 3b). Indeed,  $NAO_{cc}$  is statistically overfitted with spurious predictors, as shown by the higher validation scores for  $NAO_{mc}$ . In Figure 3c–i we verify that both ensemble reconstructions are consistent with other observational and reconstructed NAO time series (see also Extended Data Table 5). Once again,  $NAO_{mc}$  stands out, yielding generally higher correlation coefficients with the set of NAO records. We also emphasize the overall good performance of the ensemble mean, with correlations being almost systematically above the median of the ensemble. The contribution of each proxy is estimated as the total sum of its beta-weights. In both ensemble reconstructions, and most prominently in  $NAO_{mc}$ , the largest weights correspond to south and west Greenland ice cores (where the strongest impacts of the NAO have been reported<sup>15,16</sup>) and to the speleothem from Scotland (Fig. 1b, c).

To test the reliability of the  $NAO_{mc}$  reconstruction outside the observational period, we perform a perfect-model analysis of the PMIP3 simulations. We reproduce the same PCR strategy now applied to pseudo-proxies extracted from model outputs and perturbed with noise to mimic the actual proxies (Methods; ref. 25). To be more faithful to the original reconstruction, we use only the six PMIP3 simulations that represent a stable relationship of the NAO with the nine proxies in Fig. 1c. Extended Data Fig. 1 represents the correlation of the ensemble  $NAO_{mc}$  pseudo-reconstruction with the simulated NAO in three different periods: the same two subsets used before for calibration/validation, and the long validation period AD 1000–1822. As in the actual reconstructions (Fig. 3), all models depict a notably large spread for the 30-year validation scores, estimated with fewer degrees of freedom and therefore being more sensitive to stochastic noise than are the calibration statistics. In the long validation period—measuring the true skill of the pseudo-reconstruction—correlations are always significant and remain close to the range of the calibration scores. Additionally, we compute two pseudo-reconstructions:  $NAO_{Trouet}$  (using Morocco and Scotland pseudo-proxies) and  $NAO_{Lehner}$  (an improved version of  $NAO_{Trouet}$ , including hypothetical proxies from the Iberian Peninsula and Scandinavia, tested in a previous perfect-model study<sup>20</sup>). The performance of these pseudo-reconstructions (in terms of correlation with the simulated NAO) is evaluated in a 50-year moving window throughout the entire millennial runs, and compared with the performance of the mean  $NAO_{mc}$  ensemble (Extended Data Fig. 2). Our method (represented by the mean  $NAO_{mc}$  ensemble) shows clearly higher correlations with the simulated NAO ( $r_{median} = 0.44$ ) than does  $NAO_{Trouet}$  ( $r_{median} = 0.21$ ), but lower values than  $NAO_{Lehner}$  ( $r_{median} = 0.55$ ). There is therefore scope for improving our model, provided that new proxy records become available at key locations, associated with strong NAO fingerprints that are unrepresented in our data set.

Finally, we address the impact of external forcings on the yearly  $NAO_{mc}$  ensemble mean. We do not detect any significant relationship with two different total solar irradiance reconstructions<sup>26,27</sup>, even at the 11-year timescale (data not shown). In contrast, a composite analysis of the NAO response to the strongest 11 volcanic eruptions during the past millennium shows a significant ( $P < 0.05$ ) positive phase appearing two years after the selected events (Fig. 4a). This result is robust, regardless of the volcanic reconstruction considered<sup>28–30</sup> (Extended Data Table 6). One of the volcanic time series shows a second peak four years after the eruption, probably arising from misrepresentations in the timing of some events. By refining our selection (Fig. 4b)—now including 11 volcanoes represented in the three data sets, with eruption years cross-checked and corrected with historical records and information from local deposits (Extended Data Table 7)—we obtain a clearer response, only significant two years after the eruption when it

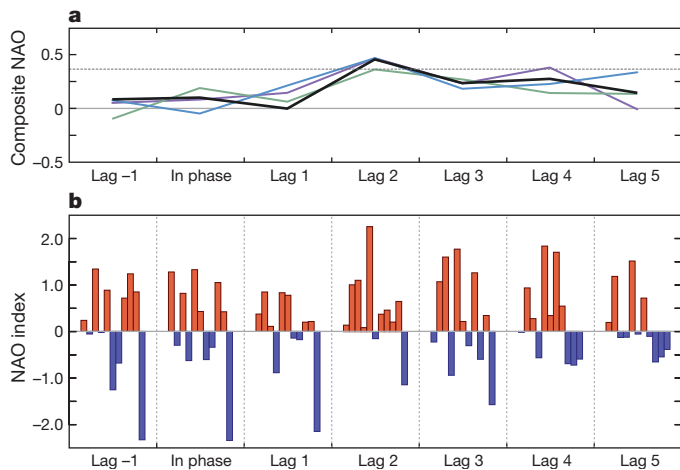


**Figure 3 | Validation of the ensemble NAO reconstructions.** **a, b**, Box-and-whisker plots of the correlations between the  $NAO_{cc}$  (yellow boxes), or  $NAO_{mc}$  (purple boxes), reconstruction ensembles and the instrumental  $NAO_{vinther}$  index, in the calibration (**a**) and validation (**b**) subsets, respectively. Small triangles indicate the value of the first significant correlation coefficient ( $P < 0.05$ ) in each ensemble. **c–i**, Correlations of the same ensemble reconstructions with other reconstructed NAO time series (Extended Data Table 4). **c–e**, NAO time series from extended sea-level-pressure data sets<sup>32–34</sup>, following the definition in ref. 23. **f–i**, NAO time series from other reconstructions<sup>7,24,35,36</sup>. Dots represent the respective correlations with the ensemble mean; these dots are filled (or unfilled) if the correlations are significant (or insignificant).

surpasses the 95% confidence level (Fig. 4a). This positive NAO response can partly explain the predominance of positive NAO phases between AD 1150 and AD 1400, a period of increased volcanic activity. However, this response to volcanoes might be overestimated in our reconstruction owing to its strong dependence on temperature-sensitive proxies from Greenland (Fig. 1c), which can respond directly to the reduced radiative conditions. Yet the fact that, in the ice cores, the relation between the registered  $\delta^{18}O$  and temperature is modulated by precipitation intermittency<sup>10</sup>, along with the lack of NAO response for the first year after the eruptions, suggests that at least part of the estimated impact of volcanoes is dynamically driven.

The identification of this volcanic influence is also consistent with observational and modelling studies<sup>8,9,31</sup>, and can be considered to be an additional validation of our reconstruction at the interannual timescale. Adding this finding to the verification via pseudo-proxies and





**Figure 4 | NAO response to large volcanic eruptions.** **a**, Composite NAO<sub>mc</sub> response to the 11 strongest eruptions in three volcanic forcing reconstructions (purple line, ref. 29; blue line, ref. 28; green line, ref. 30; Extended Data Table 6), and to a refined selection of 11 intense volcanoes common to the three data sets (black line) (see Methods). Significance is assessed following a Monte Carlo approach with 1,000 random selections of 11 years from the NAO<sub>mc</sub> mean reconstruction. The horizontal lines delimit the thresholds of significance at 95%. The 'lag' refers to the number of years after the volcanic eruption. **b**, NAO values preceding and following the onset of the 11 intense eruptions common to the three reconstructions (Extended Data Table 7).

earlier reconstructions, we conclude that the NAO<sub>mc</sub> ensemble presents a robust, highly resolved estimate of NAO variability in the last millennium, providing new insights into Northern Hemisphere climatic and environmental changes during this period, and ruling out the persistence of positive phases during the MCA. Proxy records are commonly used as a benchmark with which to assess the ability of climate models to reproduce past variability. Here, we have used reanalyses and state-of-the-art climate simulations to improve the reliability of proxy reconstructions. We have demonstrated the added value of constraining the selection of proxies with reanalysis and model outputs, and encourage the use of this approach in future reconstructions.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 29 August 2014; accepted 29 April 2015.

- Hurrell, J. W. Decadal trends in the North Atlantic Oscillation: regional temperatures and precipitation. *Science* **269**, 676–679 (1995).
- Pinto, J. G., Karremann, M. K., Born, K., Della-Marta, P. M. & Klawa, M. Loss potentials associated with European windstorms under future climate conditions. *Clim. Res.* **54**, 1–20 (2012).
- Pozo-Vazquez, D., Santos-Alamillos, F., Lara-Fanego, V., Ruiz-Arias, J., Tovar-Pescador, J. & n... in *Hydrological, Socioeconomic and Ecological Impacts of the North Atlantic Oscillation in the Mediterranean Region* Vol. 46 *Advances in Global Change Research* (eds Vicente-Serrano, S. M. & Trigo, R. M.) Ch. 15 (Springer, 2011).
- Kettlewell, P. S., Sothorn, R. B. & Koukkari, W. L. UK wheat quality and economic value are dependent on the North Atlantic Oscillation. *J. Cereal Sci.* **29**, 205–209 (1999).
- López-Moreno, J. I., Beguería, S., Vicente-Serrano, S. M. & García-Ruiz, J. M. Influence of the North Atlantic Oscillation on water resources in central Iberia: Precipitation, streamflow anomalies, and reservoir management strategies. *Wat. Resour. Res.* **43**, W09411 (2007).
- Scaife, A. A. et al. Skillful long range prediction of European and North American winters. *Geophys. Res. Lett.* **41**, 2514–2519 (2014).
- Trouet, V. et al. Persistent positive North Atlantic Oscillation mode dominated the medieval climate anomaly. *Science* **324**, 78–80 (2009).
- Graf, H. F., Kirchner, I., Robock, A. & Schult, I. Pinatubo eruption winter climate effects: model versus observations. *Clim. Dyn.* **9**, 81–93 (1993).
- Stenchikov, G. et al. Arctic Oscillation response to the 1991 Mount Pinatubo eruption: effects of volcanic aerosols and ozone depletion. *J. Geophys. Res.* **D 107**, 4803 (2002).
- Casado, M. et al. Impact of precipitation intermittency on NAO-temperature signals in proxy records. *Clim. Past* **9**, 871–886 (2013).

- Schurer, A. P., Tett, S. F. B. & Hegerl, G. C. Small influence of solar variability on climate over the past millennium. *Nature Geosci.* **7**, 104–108 (2014).
- Lamb, H. H. The early medieval warm epoch and its sequel. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **1**, 13–37 (1965).
- Olsen, J., Anderson, N. J. & Knudsen, M. F. Variability of the North Atlantic Oscillation over the past 5,200 years. *Nature Geosci.* **5**, 808–812 (2012).
- Moreno, A. et al. The Medieval Climate Anomaly in the Iberian Peninsula reconstructed from marine and lake records. *Quat. Sci. Rev.* **43**, 16–32 (2012).
- Gómez-Navarro, J. J. & Zorita, E. Atmospheric annular modes in simulations over the past millennium: no long-term response to external forcing. *Geophys. Res. Lett.* **40**, 3232–3236 (2013).
- Box, J. E. Survey of Greenland instrumental temperature records: 1873–2001. *Int. J. Climatol.* **22**, 1829–1847 (2002).
- Jones, P. D., Harpham, C. & Vinther, B. M. Winter-responding proxy temperature reconstructions and the North Atlantic Oscillation. *J. Geophys. Res.* **D 119**, 6497–6505 (2014).
- Schmütz, C., Luterbacher, J., Gyalistras, D., Xoplaki, E. & Wanner, H. Can we trust proxy-based NAO index reconstructions? *Geophys. Res. Lett.* **27**, 1135–1138 (2000).
- Raible, C. C., Lehner, F., González-Rouco, J. F. & Fernández-Donado, L. Changing correlation structures of the Northern Hemisphere atmospheric circulation from 1000 to 2100 AD. *Clim. Past* **10**, 537–550 (2014).
- Lehner, F., Raible, C. C. & Stocker, T. F. Testing the robustness of a precipitation proxy-based North Atlantic Oscillation reconstruction. *Quat. Sci. Rev.* **45**, 85–94 (2012).
- Huth, R., Bochnicek, J. & Hejda, P. The 11-year solar cycle affects the intensity and annularity of the Arctic Oscillation. *J. Atmos. Sol. Terr. Phys.* **69**, 1095–1109 (2007).
- Chen, W. & Zhou, Q. Modulation of the Arctic Oscillation and the East Asian winter climate relationships by the 11-year solar cycle. *Adv. Atmos. Sci.* **29**, 217–226 (2012).
- Vinther, B. M., Andersen, K. K., Hansen, A. W., Schmith, T. & Jones, P. D. Improving the Gibraltar/Reykjavik NAO index. *Geophys. Res. Lett.* **30**, 2222 (2003).
- Cook, E. R., D'Arrigo, R. D. & Mann, M. E. A well-verified, multiproxy reconstruction of the winter North Atlantic Oscillation index since AD 1400. *J. Clim.* **15**, 1754–1764 (2002).
- Smerdon, J. E. Climate models as a test bed for climate reconstruction methods: pseudoproxy experiments. *WIREs. Clim. Change* **3**, 63–77 (2012).
- Crowley, T. J. Causes of climate change over the past 1000 years. *Science* **289**, 270–277 (2000).
- Vieira, L. E. A., Solanki, S. K., Krivova, N. A. & Usoskin, I. Evolution of the solar irradiance during the Holocene. *Astron. Astrophys.* **531**, A6 (2011).
- Crowley, T. J. & Unterman, M. B. Technical details concerning development of a 1200 yr proxy index for global volcanism. *Earth Syst. Sci. Data* **5**, 187–197 (2013).
- Gao, C., Robock, A. & Ammann, C. Volcanic forcing of climate over the past 1500 years: an improved ice core-based index for climate models. *J. Geophys. Res.* **113**, D23111 (2008).
- Sigl, M. et al. Insights from Antarctica on volcanic forcing during the Common Era. *Nature Clim. Change* **4**, 693–697 (2014).
- Shindell, D. T., Schmidt, G. A., Mann, M. E. & Faluvegi, G. Dynamic winter climate response to large tropical volcanic eruptions since 1600. *J. Geophys. Res.* **D 109**, D05104 (2004).
- Jones, P. D. et al. Monthly mean pressure reconstructions for Europe. *Int. J. Climatol.* **19**, 347–364 (1999).
- Luterbacher, J. et al. Reconstruction of sea level pressure fields over the Eastern North Atlantic and Europe back to 1500. *Clim. Dyn.* **18**, 545–561 (2002).
- Küttel, M. et al. The importance of ship log data: reconstructing North Atlantic, European and Mediterranean sea level pressure fields back to 1750. *Clim. Dyn.* **34**, 1115–1128 (2010).
- Glueck, M. F. & Stockton, C. W. Reconstruction of the North Atlantic Oscillation, 1429–1983. *Int. J. Climatol.* **21**, 1453–1465 (2001).
- Luterbacher, J. et al. Extending North Atlantic oscillation reconstructions back to 1500. *Atmos. Sci. Lett.* **2**, 114–124 (2001).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** The work of P.O., D.S., V.M.-D. and M.C. has been financially supported by the French ANR CEPS project Green Greenland (ANR-10-CEPL-0008). F.L. and C.R. acknowledge the support of the NCCR-Climate, funded by the Swiss National Science Foundation (grant number 200020 147174). P.Y. acknowledges funding from the Swedish Research Council (grant number C0629701). We thank the paleoclimate community, in particular the PAGES2k Network, and all the institutions involved in the CMIP5/PMIP3 initiative for making available proxy data and model outputs. We also thank F. González-Rouco and M. Montoya for discussions at an earlier stage of the analysis; M. Khodri for feedback on the volcanic reconstructions; and E. Hawkins for help and advice on the statistical analysis.

**Author Contributions** P.O. gathered and prepared the data with the help of M.C., and developed the ensemble reconstruction and part of the statistical analyses. F.L. prepared the pseudo-proxies from reanalyses and models and carried out the perfect-model study. D.S. analysed the impact of external forcings on the NAO. All authors helped in conceiving the study, discussing ideas and writing the paper.

**Author Information** Data will be made available at <http://www.ncdc.noaa.gov/data-access/paleoclimatology-data>. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.O. (pablo.ortega@lscce.jpl.fr).

## METHODS

**Analysis of significance.** In the initial proxy selection of Fig. 1a, proxies are required to significantly correlate with  $\text{NAO}_{\text{vinther}}$  at the 90% confidence level. This threshold is not particularly restrictive and allows the identification of a large number of candidate time series that are potentially linked to the NAO. Since the final selection of proxies is later constrained during the calibration process, and also with reanalysis and model outputs, the risk of including time series with a spurious relationship with the NAO is small. All subsequent correlations are assessed by correcting the sample size to take into account the autocorrelation. The confidence level is kept at 90% when proxies are evaluated and in the different PCR steps, and increased to 95% for the estimation of the validation/calibration scores, and when comparing the reconstruction against other data sets. For the composite analysis of volcanic events (Fig. 4), significance is assessed at the 95% confidence level, following a Monte Carlo approach with 1,000 random equivalent selections. A confidence level of 95% is also applied for a two-tailed Student's  $t$ -test in assessing when a 100-year running average of the NAO ensemble mean is different from zero (horizontal bars in Fig. 2).

**Selection of proxies.** The initial set of proxy records (Extended Data Table 1) is the result of an extensive search throughout different palaeoclimate data repositories, such as the International Tree-Ring Data Bank (ITRDB; <https://www.ncdc.noaa.gov/data-access/paleoclimatology-data/datasets/tree-ring>), <http://www.ncdc.noaa.gov/data-access/paleoclimatology-data> and <http://www.pangaea.de/>. Many proxies (>150) from the Northern Hemisphere were found to meet the two imposed criteria: annual resolution and continuous time coverage from the eleventh to the twentieth century. These proxies were further screened for their correlation with  $\text{NAO}_{\text{vinther}}$  at the 90% confidence level. Only the 48 proxies used in our analysis met this condition.

**Climatic interpretation of proxies.** Both the seasonality and the climate variable associated with each proxy were extracted either from the original references detailed in Extended Data Table 1, or from subsequent references making use of them. For instance, a few tree-ring proxy records (such as Lily Lake) are interpreted herein in terms of annual temperature as they have already contributed to several multi-proxy reconstructions for past global and continental temperature<sup>37,38</sup>. When two possible climatic interpretations of the same proxy record exist, we kept the one leading to the highest correlations with the NAO in the twentieth century reanalysis (for example, Mammoth Peak as winter precipitation instead of summer temperature).

**Potential biases in the representation of low-frequency variability.** No post-processing has been applied to any of the proxy time series considered. All chronologies therefore correspond to the source data as they were originally provided. Note that the tree-ring records might lack some part of their low-frequency variability owing to the age-trend removal techniques. The degree of variance sacrificed may vary substantially from one record to another, because some research groups privilege more conservative detrending methods, and others prefer to retain as much low-frequency variability as possible. Ice-core records are also generally corrected for postdepositional effects, such as firn diffusion, or ice flow in the Greenland ice sheet. Firn diffusion affects only subannual variability, and may introduce biases in the seasonally resolved records. It does not affect low-frequency variability or trends. Ice flow is important only in regions of significant ice-flow rate, such as the DYE-3 drill site, and requires flow models to be corrected; it might affect to some degree the realism of the long-term trends. Other potential low-frequency biases relate to bandwidth speleothem measurements and varved lake sediments that need to be corrected for nonhomogenous growth rates and compactness, respectively. More generally, all proxy records are affected to some extent by local environmental changes (such as wind scouring for ice cores), which can perturb the recorded climatic signal. This is why it is important to perform multi-proxy reconstructions—so that most of these potential biases can be cancelled out, and the common climate signal better identified.

**PCR calculation.** The reconstruction method is based on a previous study<sup>23</sup> that uses a nested PCR to gradually adjust to the changing number of source proxy records available in consecutive periods of 25 years. Here, we consider only proxies that extend with no interruption from the eleventh to the twentieth century, so only one PCR needs to be calculated (for each ensemble realization). The PCR approach is applied as follows: all selected proxy records are first standardized with respect to the common period AD 1823–1969, and used to define a matrix (in time and space) for the period common to all proxies. This period can change depending on the actual selection, the shortest being from AD 1073 to AD 1969 when the proxy record from Firth River is available. For each ensemble realization, the corresponding subset of years previously set for calibration is extracted from that matrix. A principal-components analysis (PCA) of this reduced matrix is performed afterwards. Preisendorfer's rule  $N^{39}$  is then applied to discard those principal components (PCs) whose explained variability is undistinguishable (at the 95% significance level) from that of a matrix of red-noise processes composed of

the first autoregressive models, AR(1), of the original proxies. For this assessment, 300 independent gaussian white-noise matrices are generated (with the same dimensions as the reduced one), and a PCA is applied to each of them. In each realization, the PCs are organized in decreasing order of their variance across the Gaussian set. We then take all of the first PCs, and establish the 95th percentile of the associated variances. This step is repeated from the second to the last PC. The 95th-percentile values thus obtained are used as significance thresholds for the truncation of the original PCs. The aim of this truncation is to retain all the modes with relevant climate information, and to use them to produce a multivariate linear regression model of the NAO. To extend the reconstruction back to the eleventh century, the complete matrix of predictors is projected into the eigenvectors of the selected PCs (also known as empirical orthogonal functions, EOFs), fitting by least squares. These projections describe the variability associated with each EOF over the whole time extent of the matrix of predictors, and coincide with the original PCs in the calibration period. These are the series feeding the multivariate model to produce the actual NAO reconstruction. To correct for the loss in variance inherent to the multiple regression, the final NAO time series is restandardized with respect to the calibration period. No trends are removed before the PCR analysis, since they are also essential to reproduce the past NAO variability. This is a standard reconstruction practice<sup>24,32</sup>.

**Proxy contribution to the ensemble reconstruction.** In each individual PCR reconstruction, the beta-weights of the proxies are obtained as in ref. 24: we multiply the vector of the standardized regression coefficients (also known as beta-weights on the PC space) by the matrix of the retained eigenvectors. In practice, for an individual proxy record  $i$ , this corresponds to:

$$\beta(i) = \sum_{j=1}^n c_j \cdot \omega_j(i)$$

where  $\beta(i)$  denotes the beta-weight of the given proxy,  $n$  the number of predictors (that is, PCs) participating in the multivariate regression,  $c_j$  the regression coefficient for the predictor  $j$ , and  $\omega_j$  the loading associated with the proxy  $i$  in the eigenvector of predictor  $j$ . The beta-weight of the proxy  $i$  in the ensemble mean reconstruction is obtained as the sum of its respective beta-weights in the individual realizations.

**Estimation of reconstruction uncertainties.** Our method accounts only for regression-based uncertainties associated with the residuals in the calibration period. These are represented by the s.e. of the regression, calculated as the square root of the sum of the squared residuals divided by the degrees of freedom:

$$\text{s.e.} = \sqrt{\frac{\sum_{t=1}^n \varepsilon_t^2}{(n-2)}} = \sqrt{\frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{(n-2)}}$$

where  $n$  represents the number of time steps in the calibration period,  $\varepsilon$  the residuals,  $y$  the predicted, and  $\hat{y}$  the reconstructed variable.

Standard errors are calculated for each of the 100 selected individual reconstructions. In each of them, the uncertainty band is set at  $\pm 2$  s.e. The envelope of these bands across the different ensemble members is included in Fig. 2 to describe the total uncertainty range of each ensemble reconstruction. The perfect-model analysis supports the validity of this approach, with only about 2% of years falling outside the  $\pm 2$  s.e. envelope. The characterization of other sources of uncertainty (such as chronological errors) lies beyond the scope of our analysis.

**Sensitivity to the reconstruction choices.** Several alternative ensemble reconstructions have been performed to test the sensitivity to some of the choices in our approach. Overall, the reconstructions and the main findings remain coherent. For instance, increasing the validation period to 40 years (and reducing the calibration subsequently to 107) barely affects the final  $\text{NAO}_{\text{mc}}$  ensemble mean reconstruction. Similar results also hold if Preisendorfer's rule-N is based on a white noise floor assumption, or if proxy selection is constrained only through reanalyses or PMIP3 simulations. Additional criteria for ensemble generation, such as subsampling the proxies depending on their kind, are found to increase the uncertainty bars because of the poor skill of some members, but have a minor effect on the variability described by the mean ensemble reconstruction.

**Model-constrained ensemble reconstructions.** Given some known deficiencies in the reanalyses in terms of accurately representing variables such as precipitation in high-elevation regions (for example, Greenland or the Alps), and considering that reanalyses have different spatial resolutions and cover different time periods (Extended Data Table 2) with different degrees of confidence, we apply a relaxed constraint, so that any proxy with a significant relationship with the NAO in any of the four reanalyses is kept. Similarly, since not all PMIP3 models are equally realistic in the representation of the NAO fingerprints on surface climate<sup>40</sup>, but keeping in mind that with them we test the robustness and stability of the teleconnections, the ensemble  $\text{NAO}_{\text{mc}}$  reconstruction is additionally constrained to

those proxy records that correlate significantly with the NAO in at least half of the eight PMIP3 simulations.

**Perfect-model approach.** This technique is used to test the adequacy or realism of different assumptions or methods. It is based on the use of model outputs as substitute climate records that can be used, for instance, to assess the performance of different reconstruction strategies<sup>41,42</sup>, or the potential of decadal predictions (both statistical<sup>43</sup> and dynamical<sup>44,45</sup>). The model is assumed to be 'perfect' in the sense that it describes an idealized surrogate reality, where climate variations are physically consistent, without any measurement uncertainty, and are accessible everywhere without restrictions and for the whole extent of the simulations. In our analysis, we mimic the original proxies with model outputs (see below), and use these pseudo-proxies to reproduce the different steps of our reconstruction approach, but within the model world. One of the main advantages that this perfect model framework offers is that validation scores can also be estimated outside of the observational period AD 1823–1969. This is particularly interesting, since it allows a validation over a long time frame (AD 1000–1822), and thus provides a test of the stability of the assumed teleconnections.

**Pseudo-proxy definition.** All pseudo-proxies are defined as a  $2 \times 2$  average of the related climate variable centred on the proxy location, using the same reported seasonal window (Extended Data Table 1). These are the series used to identify the proxies that are consistent with the reanalyses and the PMIP3 simulations (Fig. 1c). The pure simulated climatic signal is later perturbed with noise for the perfect-model analysis. The noise is estimated as an AR(1) process that describes the residual of subtracting the normalized real proxy from normalized observational data at the same location. Temperature data are retrieved from Berkeley Earth (<http://berkeleyearth.org/data>) and precipitation from the Climate Research Unit (<http://badc.nerc.ac.uk/browse/badc/cru>) data sets. Our approach therefore mimics the actual noise that arises from relating a proxy signal to a single climate variable. Ten alternative pseudo-proxies are generated for ten different realizations of the noise (Extended Data Fig. 3). For the perfect-model analysis, we use the pseudo-proxy whose correlation with the simulated NAO is closer to that of the actual proxy with NAO<sub>Vinther</sub>. This is done to be more representative of the real proxy sensitivity. The hypothetical pseudo-proxies from Spain and Norway (where no real proxy is yet available), which are necessary for the NAO<sub>Lehner</sub> pseudo-reconstruction, were generated by adding Gaussian noise with properties similar to those of the twentieth century reanalysis at these locations.

**Identification of strong volcanic eruptions.** This analysis uses the three latest published reconstructions of past volcanic activity covering the past millennium. Two of them<sup>28,29</sup> combine volcanic-related information from Greenland and Antarctic ice cores, although they differ in the source records and the calibration strategy. The third reconstruction<sup>30</sup> is based exclusively on an extensive array of Antarctic ice cores, most of them not included in the former reconstructions. The combination of multiple ice-core records is an important step, accounting for age-scale and deposition uncertainties. The three reconstructions show important discrepancies regarding the timing and relative magnitude of some eruptions. This is illustrated through the selection of the 11 strongest events in each volcanic reconstruction (Extended Data Table 6). We first identify the strongest volcano according to the related reconstructed variable, and screen out all dates that are 3 years before or after the eruption (to avoid double selections). This process is repeated with the remaining dates until 11 events have been identified. Note that even when a specific eruption is present in the three volcanic series, the uncertainty in the year of occurrence can be as large as 2 years (for example, 1284, 1286, 1285). To reduce this timing uncertainty, and to ensure that we select only well supported volcanic eruptions, we extract the 25 largest eruptions in each volcanic chronology (not shown), and compare them to identify the episodes common to all reconstructions (within a  $\pm 3$  year range). This leads to a final selection of 11 common volcanoes (Extended Data Table 7). The final adopted dates for the composite analysis in Fig. 4 are extracted from historical sources (<http://www.volcano.si.edu/>)<sup>46</sup> or other reliable sedimentary records<sup>47</sup>. For the three unknown eruptions, for which no additional information is available, we took either the most repeated date, or the average date of the three reconstructions.

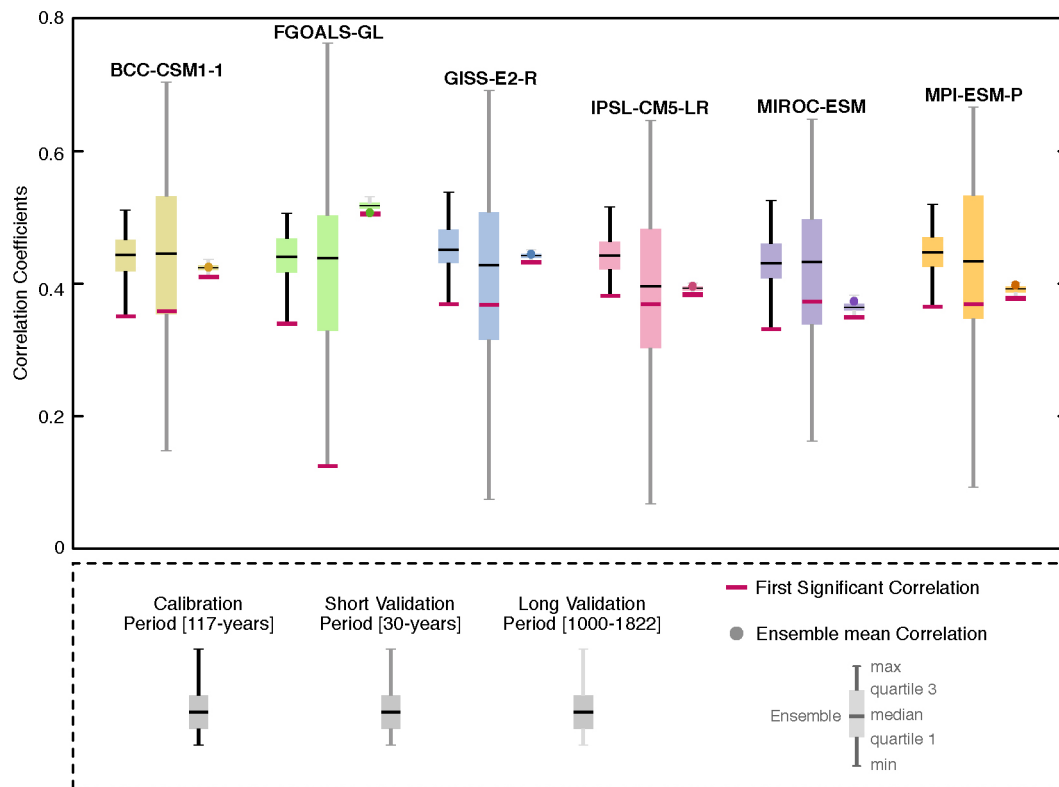
**Code availability.** The code used to produce the model-constrained ensemble reconstruction is available in the Supplementary Information.

37. Mann, M. E. *et al.* Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia. *Proc. Natl Acad. Sci. USA* **105**, 13252–13257 (2008).
38. Ahmed, M. *et al.* Continental-scale temperature variability during the past two millennia. *Nature Geosci.* **6**, 339–346 (2013).
39. Overland, J. E. & Preisendorfer, R. W. A significance test for principal components applied to a cyclone climatology. *Mon. Weath. Rev.* **110**, 1–4 (1982).
40. Davini, P. & Cagnazzo, C. On the misinterpretation of the North Atlantic Oscillation in CMIP5 models. *Clim. Dyn.* **43**, 1497–1511 (2014).

41. Baehr, J., Hirschi, J., Beismann, J. O. & Marotzke, J. Monitoring the meridional overturning circulation in the North Atlantic: a model-based array design study. *J. Mar. Res.* **62**, 283–312 (2004).
42. González-Rouco, J., Beltrami, H., Zorita, E. & von Storch, H. Simulation and inversion of borehole temperature profiles in surrogate climates: spatial distribution and surface coupling. *Geophys. Res. Lett.* **33**, L01703 (2006).
43. Hawkins, E., Robson, J., Sutton, R., Smith, D. & Keenlyside, N. Evaluating the potential for statistical decadal predictions of sea surface temperatures with a perfect model approach. *Clim. Dyn.* **37**, 2495–2509 (2011).
44. Griffies, S. M. & Bryan, K. A predictability study of simulated North Atlantic multidecadal variability. *Clim. Dyn.* **13**, 459–487 (1997).
45. Collins, M. & Sinha, B. Predictability of decadal variations in the thermohaline circulation and climate. *Geophys. Res. Lett.* **30**, 1306 (2003).
46. Siebert, L., Simkin, T. & Kimberly, P. *Volcanoes of the World* 3rd edn (Univ. California Press, 2010).
47. Lavigne, F. *et al.* Source of the great A.D. 1257 mystery eruption unveiled, Samalas volcano, Rinjani volcanic complex, Indonesia. *Proc. Natl Acad. Sci. USA* **110**, 16742–16747 (2013).
48. Miller, H. & Schwager, M. Accumulation rate and stable oxygen isotope ratios of ice core ngt14C93.2 from the North Greenland Traverse. <http://dx.doi.org/10.1594/PANGAEA.57158> (2000).
49. Andersen, K. K. *et al.* Retrieving a common accumulation record from Greenland ice cores for the past 1800 years. *J. Geophys. Res.* **111**, D15106 (2006).
50. Cuffey, K. M. *et al.* Large Arctic temperature change at the Wisconsin-Holocene glacial transition. *Science* **270**, 455–458 (1995).
51. Fisher, D. A., Koerner, R. M. & Reeh, N. Holocene climatic records from Agassiz Ice Cap, Ellesmere Island, NWT, Canada. *Holocene* **5**, 19–24 (1995).
52. Vinther, B. M. *et al.* Climatic signals in multiple highly resolved stable isotope records from Greenland. *Quat. Sci. Rev.* **29**, 522–538 (2010).
53. Grootes, P. & Stuiver, M. Oxygen 18/16 variability in Greenland snow and ice with  $10^{-3}$  to  $10^5$  year time resolution. *J. Geophys. Res.* **102**, 26455–26470 (1997).
54. Wilhelm, B. *et al.* 1400 years of extreme precipitation patterns over the Mediterranean French Alps and possible forcing mechanisms. *Quat. Res.* **78**, 1–12 (2012).
55. Moore, J. J., Hughen, K. A., Miller, G. H. & Overpeck, J. T. Little Ice Age recorded in summer temperature reconstruction from varved sediments of Donard Lake, Baffin Island, Canada. *J. Paleolimnol.* **25**, 503–517 (2001).
56. Larsen, D. J., Miller, G. H., Geirsdottir, A. & Thordarson, T. A 3000-year varved record of glacier activity and climate change from the proglacial lake Hlíðaráttvatn, Iceland. *Quat. Sci. Rev.* **30**, 2715–2731 (2011).
57. Cook, T. L., Bradley, R. S., Stoner, J. S. & Francus, P. Five thousand years of sediment transfer in a high arctic watershed recorded in annually laminated sediments from Lower Murray Lake, Ellesmere Island, Nunavut, Canada. *J. Paleolimnol.* **41**, 77–94 (2009).
58. McCabe-Glynn, S. *et al.* Variable North Pacific influence on drought in southwestern North America since AD 854. *Nature Geosci.* **6**, 617–621 (2013).
59. Proctor, C., Baker, A., Barnes, W. & Gilmour, M. A thousand year speleothem proxy record of North Atlantic climate from Scotland. *Clim. Dyn.* **16**, 815–820 (2000).
60. Seim, A. *et al.* Climate sensitivity of a millennium-long pine chronology from Albania. *Clim. Res.* **51**, 217–228 (2012).
61. Büntgen, U., Frank, D. C., Nievergelt, D. & Esper, J. Summer temperature variations in the European Alps, A.D. 755–2004. *J. Clim.* **19**, 5606–5623 (2006).
62. Stahle, D. W. *International Tree-Ring Data Bank AR050*. <https://www.ncdc.noaa.gov/data-access/paleoclimatology-data/datasets/tree-ring> (1996).
63. Stahle, D. W. & Cleaveland, M. K. *International Tree-Ring Data Bank FL001*. <https://www.ncdc.noaa.gov/data-access/paleoclimatology-data/datasets/tree-ring> (2005).
64. Briffa, K. R. *et al.* Trends in recent temperature and radial growth spanning 2000 years across northwest Eurasia. *Phil. Trans. R. Soc.* **363**, 2271–2284 (2008).
65. Tsh, R. *International Tree-Ring Data Bank CA051*. <https://www.ncdc.noaa.gov/data-access/paleoclimatology-data/datasets/tree-ring> (1994).
66. Bunn, A. G., Graumlich, L. J. & Urban, D. L. Trends in twentieth-century tree growth at high elevations in the Sierra Nevada and White Mountains, USA. *Holocene* **15**, 481–488 (2005).
67. Woodhouse, C. A. & Brown, P. M. *International Tree-Ring Data Bank CO572*. <https://www.ncdc.noaa.gov/data-access/paleoclimatology-data/datasets/tree-ring> (2006).
68. Anchukaitis, K. J. *et al.* Tree-ring-reconstructed summer temperatures from northwestern North America during the last nine centuries. *J. Clim.* **26**, 3001–3012 (2013).
69. Young, G. H. F. *et al.* Changes in atmospheric circulation and the Arctic Oscillation preserved within a millennial length reconstruction of summer cloud cover from northern Fennoscandia. *Clim. Dyn.* **39**, 495–507 (2012).
70. Büntgen, U., Frank, D., Neuenschwander, T. & Esper, J. Fading temperature sensitivity of Alpine tree growth at its Mediterranean margin and associated effects on large-scale climate reconstructions. *Clim. Change* **114**, 651–666 (2012).
71. Stahle, D. W. *International Tree-Ring Data Bank LA001*. <https://www.ncdc.noaa.gov/data-access/paleoclimatology-data/datasets/tree-ring> (1996).
72. Esper, J. *et al.* Long-term drought severity variations in Morocco. *Geophys. Res. Lett.* **34**, L17702 (2007).
73. Graumlich, L. J., Pisarcic, M. F. J., Waggoner, L. A., Littell, J. S. & King, J. C. Upper Yellowstone River flow and teleconnections with Pacific basin climate variability during the past three centuries. *Clim. Change* **59**, 245–262 (2003).
74. Touchan, R., Woodhouse, C. A., Meko, D. M. & Allen, C. Millennial precipitation reconstruction for the Jemez Mountains, New Mexico, reveals changing drought signal. *Int. J. Climatol.* **31**, 896–906 (2011).

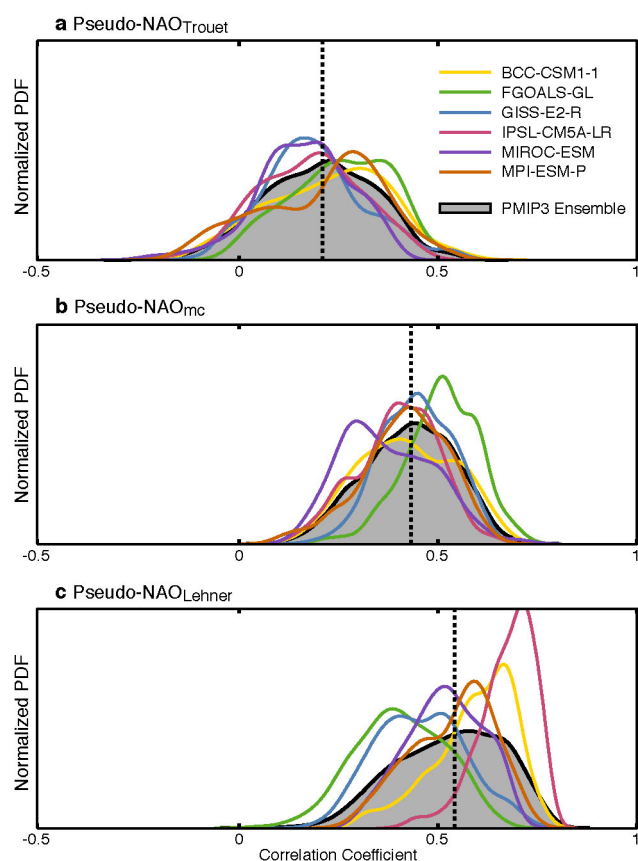


75. Esper, J. *et al.* Orbital forcing of tree-ring data. *Nature Clim. Change* **2**, 862–866 (2012).
76. Graybill, D. A. *International Tree-Ring Data Bank* NV516. <https://www.ncdc.noaa.gov/data-access/paleoclimatology-data/datasets/tree-ring> (1994).
77. Graybill, D. A. *International Tree-Ring Data Bank* NV517. <https://www.ncdc.noaa.gov/data-access/paleoclimatology-data/datasets/tree-ring> (1994).
78. Salzer, M. & Kipfmüller, K. Reconstructed temperature and precipitation on a millennial timescale from tree-rings in the Southern Colorado Plateau, USA. *Clim. Change* **70**, 465–487 (2005).
79. Stahle, D. W. & Cleaveland, M. K. Tree-ring reconstructed rainfall over the southeastern USA during the medieval warm period and the little ice age. *Clim. Change* **26**, 199–212 (1994).
80. Naurzbaev, M. M., Vaganov, E. A., Sidorova, O. V. & Schweingruber, F. H. Summer temperatures in eastern Taimyr inferred from a 2427-year late-Holocene tree-ring chronology and earlier floating series. *Holocene* **12**, 727–736 (2002).
81. Buntgen, U. *et al.* Filling the Eastern European gap in millennium-long temperature reconstructions. *Proc. Natl Acad. Sci. USA* **110**, 1773–1778 (2013).
82. Schweingruber, F. H. *International Tree-Ring Data Bank* SWIT177. <https://www.ncdc.noaa.gov/data-access/paleoclimatology-data/datasets/tree-ring>.
83. Graybill, D. A. *International Tree-Ring Data Bank* UT508. <https://www.ncdc.noaa.gov/data-access/paleoclimatology-data/datasets/tree-ring> (1994).
84. Graybill, D. A. *International Tree-Ring Data Bank* UT509. <https://www.ncdc.noaa.gov/data-access/paleoclimatology-data/datasets/tree-ring> (1994).
85. Kalnay, E. *et al.* The NCEP/NCAR 40-year reanalysis project. *Bull. Am. Meteorol. Soc.* **77**, 437–471 (1996).
86. Compo, G. *et al.* The twentieth century reanalysis project. *Q. J. R. Meteorol. Soc.* **137**, 1–28 (2011).
87. Dee, D. P. *et al.* The ERA-interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* **137**, 553–597 (2011).
88. Uppala, S. M. *et al.* The ERA-40 re-analysis. *Q. J. R. Meteorol. Soc.* **131**, 2961–3012 (2005).
89. Zhang, J. & Wu, T. W. The impact of external forcings on climate during the past millennium: Results from transient simulation with BCC\_CSM1.1. *European Geosciences Union General Assembly* abstr. EGU2012-448 (2012).
90. Landrum, L. *et al.* Last millennium climate and its variability in CCSM4. *J. Clim.* **26**, 1085–1111 (2012).
91. Phipps, S. *et al.* The CSIRO Mk3L climate system model version 1.0–Part 2: response to external forcings. *Geosci. Model Dev.* **5**, 649–682 (2012).
92. Zhou, T., Li, B., Man, W., Zhang, L. & Zhang, J. A comparison of the Medieval Warm Period, Little Ice Age and 20th century warming simulated by the FGOALS climate system model. *Chin. Sci. Bull.* **56**, 3028–3041 (2011).
93. Schmidt, G. A. *et al.* Configuration and assessment of the GISS ModelE2 contributions to the CMIP5 archive. *J. Adv. Model. Earth Syst.* **6**, 141–184 (2014).
94. Sueyoshi, T. *et al.* Set-up of the PMIP3 paleoclimate experiments conducted using an Earth system model, MIROC-ESM. *Geosci. Model Dev. Discuss.* **6**, 819–836 (2013).
95. Jungclauss, J. H., Lohmann, K. & Zanchettin, D. Enhanced 20th century heat transfer to the Arctic simulated in the context of climate variations over the last millennium. *Clim. Past* **10**, 2895–2924 (2014).



**Extended Data Figure 1 | Verification of the pseudo-reconstructions.** Box-and-whisker plots showing the correlations between the six pseudo-reconstructed NAO<sub>mc</sub> ensembles and the corresponding true simulated NAOs, in three independent subsets of years (from left to right for each ensemble): the 117 years selected for calibration (specific to each ensemble realization), the remaining 30 years that need to be validated (following the original strategy),

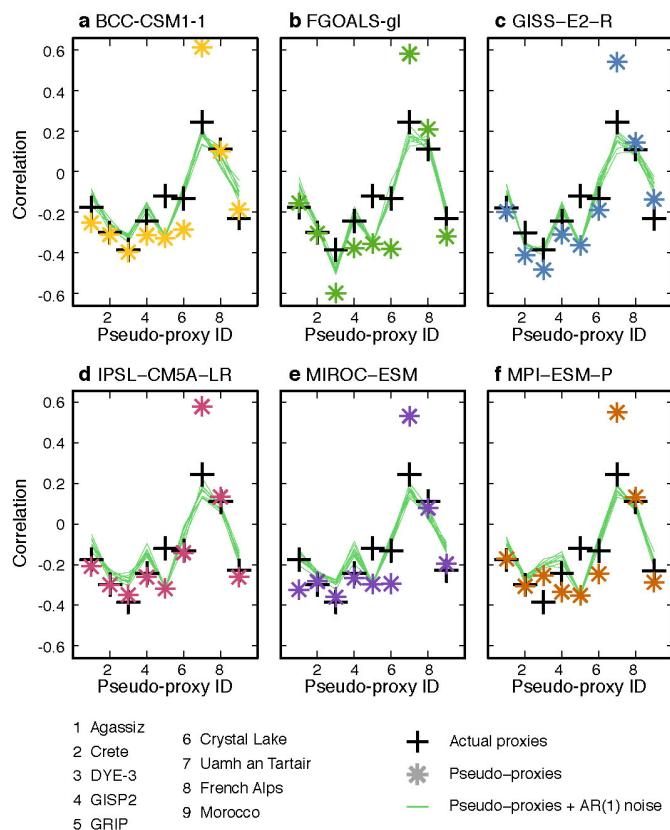
and a long validation period from AD 1000–1822 (possible thanks to the perfect-model approach). In this latter case, since the period is the same for all the ensemble realizations, the correlation is also calculated for the ensemble mean (filled coloured dots). Red horizontal lines indicate the first significant correlation coefficient ( $P < 0.05$ ) of the ensemble.



#### Extended Data Figure 2 | Comparison of three alternative reconstructions.

The figure shows probability density functions (PDFs) for the 50-year moving-window correlations between the simulated NAO and three alternative pseudo-reconstructions in the ensemble of PMIP3 simulations (Extended Data Table 3): **a**, NAO<sub>Trouet</sub>; **b**, NAO<sub>mc</sub>; **c**, NAO<sub>Lehner</sub>. Vertical dotted lines represent the median correlation for the ensemble. Each coloured line represents a different pseudo-reconstructed NAO ensemble.





### Extended Data Figure 3 | NAO-pseudo-proxy relationships.

**a–f**, Correlations between the simulated NAO and the associated pseudo-proxies in the PMIP3 runs. Pseudo-proxies are defined from a pure climatic signal (coloured asterisks), and subsequently perturbed with AR(1) noises (thin green lines). For the perturbed pseudo-proxy definitions, ten different realizations of the noise are considered. The correlations between the actual proxies and  $NAO_{Vinter}$  in the common period AD 1823–1969 are also shown for comparison (thick black crosses).

Extended Data Table 1 | Proxy description

| N  | Site  | Archive       | Proxy type                      | LON    | LAT  | Time Span | Correlation in overlap period | Related variable | Seasonality  | Refs |
|----|---|---------------|---------------------------------|--------|------|-----------|-------------------------------|------------------|--------------|------|
| 1  | B18   | Ice core      | Snow Accumulation               | -36.4  | 76.6 | 1000-1992 | 0.19                          | Precip           | Annual       | 48   |
| 2  | Crete                                       | Ice core      | Snow Accumulation               | -39.7  | 71   | 1000-1973 | -0.19                         | Precip           | Annual       | 49   |
| 3  | GISP2                                       | Ice core      | Snow Accumulation               | -40.9  | 72.5 | 1000-1988 | -0.12                         | Precip           | Annual       | 50   |
| 4  | <b>Agassiz (A79)*</b>                       | Ice core      | $\delta^{18}\text{O}^{\dagger}$ | -73.1  | 80.7 | 1000-1972 | -0.18                         | SAT              | Annual       | 51   |
| 5  | <b>Crete</b>                                | Ice core      | $\delta^{18}\text{O}^{\dagger}$ | -37.3  | 71.1 | 1000-1973 | -0.31                         | SAT              | NDJFMA       | 52   |
| 6  | <b>DYE-3 (stack)</b>                        | Ice core      | $\delta^{18}\text{O}^{\dagger}$ | -43.8  | 65.2 | 1000-1978 | -0.39                         | SAT              | NDJFMA       | 52   |
| 7  | <b>GISP2</b>                                | Ice core      | $\delta^{18}\text{O}^{\dagger}$ | -38.5  | 72.6 | 1000-1987 | -0.21                         | SAT              | Annual       | 53   |
| 8  | <b>GRIP</b>                                 | Ice core      | $\delta^{18}\text{O}^{\dagger}$ | -37.5  | 72.6 | 1000-1979 | -0.13                         | SAT              | NDJFMA       | 52   |
| 9  | Allos Lake                                  | Lake Sediment | Flood deposit thickness         | 6.7    | 44.2 | 1000-2009 | 0.11                          | Precip           | SON          | 54   |
| 10 | Donard Lake                                 | Lake Sediment | Varve thickness                 | -61.4  | 66.7 | 1000-1995 | -0.14                         | SAT              | JJA          | 55   |
| 11 | Hvítárvatn Lake                             | Lake Sediment | Varve thickness                 | -19.8  | 64.6 | 1000-2000 | -0.11                         | SAT              | JJAS         | 56   |
| 12 | Lower Murray Lake                           | Lake Sediment | Mass accumulation               | -69.5  | 81.3 | 1000-1969 | -0.18                         | SAT              | JJA          | 57   |
| 13 | <b>Crystal Cave</b>                         | Speleothem    | $\delta^{18}\text{O}^{\dagger}$ | -18.8  | 36.6 | 1000-2007 | -0.15                         | SAT              | Annual       | 58   |
| 14 | <b>Uamh an Tartair</b>                      | Speleothem    | Bandwidth                       | -4.9   | 58.1 | 1000-1995 | 0.24                          | Precip           | DJFM         | 59   |
| 15 | Balkan Peninsula                            | Tree ring     | Tree-ring width                 | 20.0   | 41.0 | 1000-2008 | -0.10                         | SAT              | February     | 60   |
| 16 | European Alps                               | Tree ring     | Tree-ring MXD                   | 7.5    | 46.0 | 1000-2004 | 0.11                          | SAT              | JJAS         | 61   |
| 17 | Black Swamp                                 | Tree ring     | Tree-ring width                 | -91.3  | 35.2 | 1019-1980 | 0.16                          | SAT              | Annual       | 62   |
| 18 | Mayberry Slough                             | Tree ring     | Tree-ring width                 | -91.3  | 35.6 | 1000-1990 | 0.12                          | SAT <sup>‡</sup> | Annual       | 63   |
| 19 | Avam-Taimyr                                 | Tree ring     | Tree-ring width                 | 101.0  | 72.0 | 1000-2000 | -0.10                         | SAT              | July         | 64   |
| 20 | San Gorgonio                                | Tree ring     | Tree-ring width                 | -116.8 | 34.1 | 1000-1970 | 0.19                          | SAT              | Annual       | 65   |
| 21 | Flower Lake                                 | Tree ring     | Tree-ring width                 | -118.4 | 36.8 | 1000-1987 | -0.14                         | Precip           | NDJFM        | 66   |
| 22 | Timber Gap Upper                            | Tree ring     | Tree-ring width                 | -118.6 | 36.5 | 1000-1987 | -0.13                         | Precip           | NDJFM        | 66   |
| 23 | Cirque Peak                                 | Tree ring     | Tree-ring width                 | -118.2 | 36.5 | 1000-1987 | -0.13                         | Precip           | NDJFM        | 66   |
| 24 | Mammoth Peak                                | Tree ring     | Tree-ring width                 | -119.3 | 37.9 | 1000-1996 | -0.10                         | Precip           | NDJFM        | 66   |
| 25 | Boreal Plateau                              | Tree ring     | Tree-ring width                 | -118.6 | 36.5 | 1000-1992 | -0.14                         | Precip           | NDJFM        | 66   |
| 26 | Upper Wright Lakes                          | Tree ring     | Tree-ring width                 | -118.4 | 36.6 | 1000-1992 | -0.14                         | Precip           | NDJFM        | 66   |
| 27 | Hamilton                                    | Tree ring     | Tree-ring width                 | -118.9 | 37.0 | 1000-1988 | -0.22                         | Precip           | NDJFM        | 66   |
| 28 | Lily Lake                                   | Tree ring     | Tree-ring width                 | -105.6 | 40.3 | 1000-1998 | -0.12                         | SAT <sup>‡</sup> | Annual       | 67   |
| 29 | Firth River                                 | Tree ring     | Tree-ring MXD                   | -141.6 | 68.7 | 1073-2002 | 0.14                          | SAT              | JJA          | 68   |
| 30 | Choctawhatchee River                        | Tree ring     | Tree-ring width                 | -85.9  | 30.5 | 1000-1992 | 0.14                          | SAT <sup>‡</sup> | Annual       | 63   |
| 31 | Forfjordalen                                | Tree ring     | Tree-ring $\delta^{13}\text{C}$ | 15.7   | 68.8 | 1000-2001 | -0.14                         | Cloud%           | JJA          | 69   |
| 32 | <b>French Alps</b>                          | Tree ring     | Tree-ring width                 | 7.5    | 44.0 | 1000-2007 | 0.10                          | SAT              | AMJJA        | 70   |
| 33 | Big Cypress                                 | Tree ring     | Tree-ring width                 | -93.0  | 32.3 | 1000-1988 | 0.10                          | SAT <sup>‡</sup> | Annual       | 71   |
| 34 | <b>Morocco</b>                              | Tree ring     | Tree-ring width                 | -5.0   | 33.8 | 1049-2001 | -0.24                         | SPI              | FMAMJ        | 72   |
| 35 | Yellow Mountain Ridge I                     | Tree ring     | Tree-ring width                 | -111.3 | 45.3 | 1000-1998 | -0.13                         | Precip           | NDJFM        | 73   |
| 36 | Yellow Mountain Ridge I (Entire Bark Trees) | Tree ring     | Tree-ring width                 | -111.3 | 45.3 | 1000-1998 | -0.11                         | Precip           | NDJFM        | 73   |
| 37 | Mesa Alta                                   | Tree ring     | Tree-ring width                 | -106.6 | 36.2 | 1000-2007 | 0.13                          | Precip           | ONDJFMAMJ    | 74   |
| 38 | Finland                                     | Tree ring     | Tree-ring MXD                   | 25.0   | 68.0 | 1000-2006 | -0.11                         | SAT              | JJA          | 75   |
| 39 | Hill 10842                                  | Tree ring     | Tree-ring width                 | -114.2 | 38.9 | 1000-1984 | 0.20                          | SAT <sup>‡</sup> | Annual       | 76   |
| 40 | Springs Mountains Lower                     | Tree ring     | Tree-ring width                 | -115.7 | 36.3 | 1000-1984 | 0.11                          | SAT              | Annual       | 77   |
| 41 | S. Colorado Plateau I                       | Tree ring     | Tree-ring width                 | -110.0 | 37.0 | 1000-1987 | 0.18                          | Precip           | October-July | 78   |
| 42 | S. Colorado Plateau II                      | Tree ring     | Tree-ring width                 | -110.0 | 37.0 | 1000-1996 | -0.11                         | SAT              | Annual       | 78   |
| 43 | Four Holes Swamp                            | Tree ring     | Tree-ring width                 | -80.4  | 33.2 | 1001-1985 | -0.10                         | Precip           | MAMJ         | 79   |
| 44 | Taimyr - Putoran                            | Tree ring     | Tree-ring width                 | 103.0  | 71.3 | 1000-1996 | -0.11                         | SAT              | Annual       | 80   |
| 45 | Tatra Region                                | Tree ring     | Tree-ring width                 | 20.0   | 49.0 | 1040-2011 | -0.12                         | SAT              | May-June     | 81   |
| 46 | Lauenen+ div. Stao                          | Tree ring     | Tree-ring width                 | 7.2    | 46.4 | 1000-1976 | 0.14                          | SAT              | JJA          | 82   |
| 47 | Wild Horse Ridge                            | Tree ring     | Tree-ring width                 | -111.1 | 39.4 | 1000-1985 | 0.19                          | SAT <sup>‡</sup> | Annual       | 83   |
| 48 | Mammoth Creek                               | Tree ring     | Tree-ring width                 | -112.7 | 37.7 | 1000-1989 | 0.18                          | SAT <sup>‡</sup> | Annual       | 84   |

\* Proxies participating in the  $\text{NAO}_{\text{m}}$  reconstruction are highlighted in bold.

<sup>†</sup> Isotope values are referred to Vienna Standard Mean Ocean Water (VSMOW).

<sup>‡</sup> These proxy records are sensitive to negative changes in the related variable.

Data are from refs 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84. LON, longitude; LAT, latitude; SAT, surface air temperature.

**Extended Data Table 2 | Description of the reanalyses.**

| Reanalysis                      | Time Interval Analysed | Horizontal Resolution                 | Reference |
|---------------------------------|------------------------|---------------------------------------|-----------|
| NCEP/NCAR Reanalysis 1          | 1949-2008              | Global grid (2.5° x 2.5)              | Ref. 85   |
| Twentieth Century Reanalysis V2 | 1872-2010              | T62 Gaussian grid (~2.0° x 2.0°)      | Ref. 86   |
| ERA Interim                     | 1980-2011              | T255 Gaussian grid (~0.75° x 0.75°)   | Ref. 87   |
| ERA40                           | 1959-2001              | T159 Gaussian grid (~1.125° x 1.125°) | Ref. 88   |

Data are from refs 85, 86, 87, 88. NCEP, National Centers for Environmental Prediction; NCAR, the National Center for Atmospheric Research.



**Extended Data Table 3 | Description of the simulations.**

| Simulation*    | PMIP3/CMIP5 family of experiments   | Time Span    | Reference            |
|----------------|-------------------------------------|--------------|----------------------|
| BCC-CSM1-1     | past1000(r1i1p1)+historical(r1i1p1) | 850-2005 CE  | Ref. 89              |
| CCSM4          | past1000(r1i1p1)+continuation       | 850-2005 CE  | Ref. 90              |
| CSIRO-MK3L-1-2 | past1000(r1i1p1)+continuation       | 851-2005 CE  | Ref. 91 <sup>†</sup> |
| FGOALS-g1      | past1000(r1i1p1)                    | 1000-1999 CE | Ref. 92              |
| GISS-E2-R      | past1000(r1i1p121)+continuation     | 850-2005 CE  | Ref. 93              |
| IPSL-CM5A-LR   | past1000(r1i1p1)+continuation       | 850-1999 CE  |                      |
| MIROC-ESM      | past1000(r1i1p1)+continuation       | 850-2005 CE  | Ref. 94              |
| MPI-ESM-P      | past1000(r1i1p1)+continuation       | 850-2005 CE  | Ref. 95              |

\* Further information on the resolution and forcings used in the past 1,000 simulations can be found at [https://wiki.lsce.ipsl.fr/pmip3/doku.php/pmip3:database:status#past1000\\_experiment\\_status\\_and\\_bc](https://wiki.lsce.ipsl.fr/pmip3/doku.php/pmip3:database:status#past1000_experiment_status_and_bc)

<sup>†</sup>This article describes a different simulation, but specifies the differences with respect to the PMIP3 run.

Data are from refs 89, 90, 91, 92, 93, 94, 95.

**Extended Data Table 4 | Principal-component (PC) contributions to the ensemble reconstructions.**

| Simulation | PC1 | PC2 | PC3 | PC4 | PC5 |
|------------|-----|-----|-----|-----|-----|
| NAOcc      | 100 | 98  | 49  | 11  | 2   |
| NAOmc      | 100 | 0   | 0   | 0   | 0   |

**Extended Data Table 5 | Description of the NAO/sea-level pressure (SLP) records used for validation.**

| Reference | Data sources*                            | Regions represented                    | Validation period |
|-----------|--|--|-------------------|
| Ref. 32   | Instrumental pressure data               | Eastern North Atlantic/ Europe         | 1781-1822 CE      |
| Ref. 33   | Instrumental, documentary and proxy data | Eastern North Atlantic/ Europe         | 1659-1822 CE      |
| Ref. 34   | Instrumental SLP and wind from ship logs | Eastern North Atlantic/ Europe         | 1751-1822 CE      |
| Ref. 35   | 5 tree-ring and 2 ice cores              | Morocco/Finland/Greenland              | 1429-1822 CE      |
| Ref. 36   | Instrumental, documentary and proxy data | Eastern North Atlantic/ Europe         | 1500-1822 CE      |
| Ref. 24   | 367 records (tree-rings and ice cores)   | Europe/Greenland/Eastern North America | 1400-1822 CE      |
| Ref. 7    | 1 speleothem and 1 tree-ring chronology  | Scotland/Morocco                       | 1049-1822 CE      |

\* Only the two proxies used in ref. 7. have been included in our analysis.

Data are from refs 7, 24, 32, 33, 34, 35, 36. CE, Common Era, equivalent to AD.



**Extended Data Table 6 | The 11 largest volcanic eruptions between AD 1049 and AD 1969 in three alternative reconstructions.**

| Reconstruction | Reconstructed variable  | List of volcanoes                                      |
|----------------|---|--|
| Ref. 28        | Global aerosol optical depth at 550 nm*                                   | 1229,1258,1286,1456,1600,1641,1674,1696,1809,1816,1884 |
| Ref. 29        | Global total stratospheric sulfate aerosol injection (in Tg) <sup>†</sup> | 1167,1227,1258,1275,1284,1452,1600,1641,1783,1809,1815 |
| Ref. 30        | Antarctic volcanic sulfate (in kg/km <sup>2</sup> ) <sup>‡</sup>          | 1169,1229,1257,1276,1285,1344,1458,1600,1694,1809,1815 |

\* Data from [http://www1.ncdc.noaa.gov/pub/data/paleo/climate\\_forcing/volcanic\\_aerosols/crowley2013/crowley2013aod-reff.txt](http://www1.ncdc.noaa.gov/pub/data/paleo/climate_forcing/volcanic_aerosols/crowley2013/crowley2013aod-reff.txt)

<sup>†</sup> Revised version from [http://climate.envsci.rutgers.edu/IVI2/IVI2TotalInjection\\_501-2000Version2.txt](http://climate.envsci.rutgers.edu/IVI2/IVI2TotalInjection_501-2000Version2.txt)

<sup>‡</sup> Data available in the supplement of ref. 30.

**Extended Data Table 7 | Eleven large volcanic eruptions common to the three alternative reconstructions.**

| Volcano      | Country     | Final selected date     | Date from Ref. 28 | Date from Ref. 29 | Date from Ref. 30 |
|--------------|-------------|-------------------------|-------------------|-------------------|-------------------|
| Unknown      |             | <b>1229</b>             | 1229              | 1227              | 1229              |
| Samalas      | Indonesia   | <b>1257*</b>            | 1258              | 1258              | 1257              |
| Unknown      |             | <b>1285</b>             | 1286              | 1284              | 1285              |
| Huaynaputina | Peru        | <b>1600<sup>†</sup></b> | 1600              | 1600              | 1600              |
| Parker       | Philippines | <b>1640<sup>†</sup></b> | 1641              | 1641              | 1641              |
| Serua        | Indonesia   | <b>1693<sup>†</sup></b> | 1696              | 1693              | 1694              |
| Unknown      |             | <b>1809</b>             | 1809              | 1809              | 1809              |
| Tambora      | Indonesia   | <b>1815<sup>†</sup></b> | 1816              | 1815              | 1815              |
| Cosiguina    | Nicaragua   | <b>1835<sup>†</sup></b> | 1835              | 1835              | 1834              |
| Krakatau     | Indonesia   | <b>1883<sup>†</sup></b> | 1884              | 1883              | 1884              |
| Agung        | Indonesia   | <b>1963<sup>†</sup></b> | 1964              | 1963              | 1963              |

\* Date recently constrained using local deposits and historical records<sup>47</sup>.

<sup>†</sup> Date extracted from historical observations<sup>46</sup>.

# Hallucigenia's head and the pharyngeal armature of early ecdysozoans

Martin R. Smith<sup>1</sup> & Jean-Bernard Caron<sup>2,3</sup>

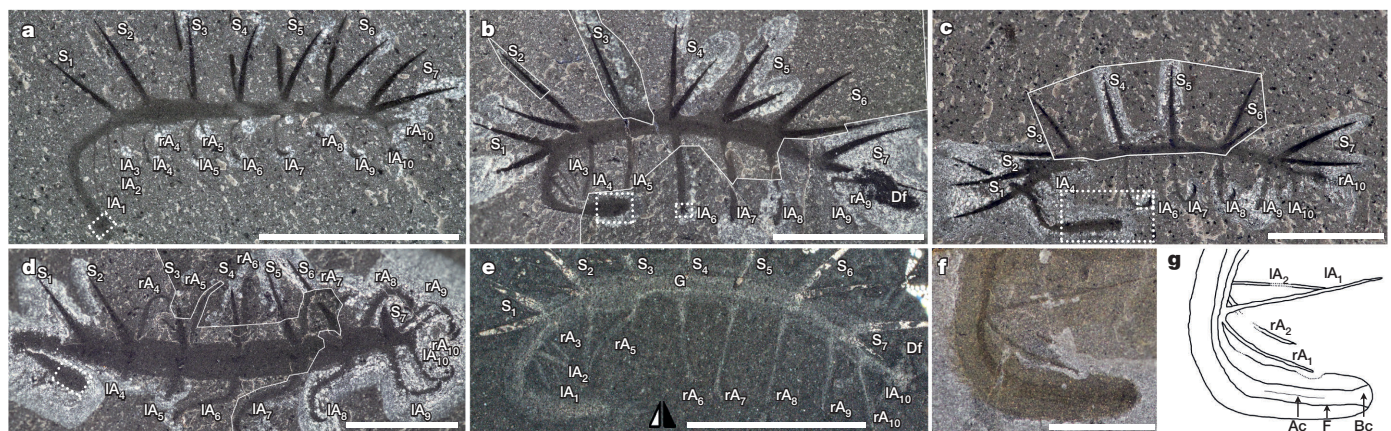
The molecularly defined clade Ecdysozoa<sup>1</sup> comprises the panarthropods (Euarthropoda, Onychophora and Tardigrada) and the cycloneurialian worms (Nematoda, Nematomorpha, Priapulida, Loricifera and Kinorhyncha). These disparate phyla are united by their means of moulting, but otherwise share few morphological characters—none of which has a meaningful fossilization potential. As such, the early evolutionary history of the group as a whole is largely uncharted. Here we redescribe the 508-million-year-old stem-group onychophoran *Hallucigenia sparsa*<sup>2–6</sup> from the mid-Cambrian Burgess Shale. We document an elongate head with a pair of simple eyes, a terminal buccal chamber containing a radial array of sclerotized elements, and a differentiated foregut that is lined with acicular teeth. The radial elements and pharyngeal teeth resemble the sclerotized circumoral elements and pharyngeal teeth expressed in tardigrades<sup>7–9</sup>, stem-group euarthropods<sup>10–12</sup> and cycloneurialian worms<sup>13</sup>. Phylogenetic results indicate that equivalent structures characterized the ancestral panarthropod and, seemingly, the ancestral ecdysozoan, demonstrating the deep homology of panarthropod and cycloneurialian mouthparts, and providing an anatomical synapomorphy for the ecdysozoan supergroup.

Although Cambrian ecdysozoans offer an unrivalled perspective on early ecdysozoan evolution<sup>6,14</sup>, considerable uncertainty surrounds the morphology of the ancestral ecdysozoan. One of the few areas of agreement is that this ancestor bore a pharynx lined with ectodermally derived, periodically moulted cuticle<sup>7</sup> and opening at a terminal mouth<sup>15</sup>.

In many ecdysozoan taxa, the pharynx is lined with sclerotized teeth<sup>9,10,12,13,16</sup>, and the mouth is surrounded by circumoral elements.

The typical cycloneurialian mouth is surrounded by a ring of spines<sup>13</sup>; the tardigrade mouth bears circumoral lamellae<sup>11,14,17</sup>; stem-group euarthropods (such as *Hurdia*, *Kerygmachela* and *Jianshanopodia*) exhibit various lamellae and plates<sup>10–12</sup>; and the onychophoran mouth is enclosed by pustular lips. These elements have formerly been regarded as homologous throughout Ecdysozoa<sup>12,15,18–21</sup>. However, the non-sclerotized lips of onychophorans are not strictly circumoral<sup>22</sup>, and onychophorans conspicuously lack pharyngeal teeth<sup>16</sup>. This suggests two possibilities: (1) a foregut armature of circumoral elements and pharyngeal teeth did exist in the ancestral ecdysozoan, but was secondarily lost in onychophorans; or (2) homoplasious armatures arose independently in Panarthropoda (either once or twice, depending on panarthropod relationships<sup>6,23</sup>) and Cycloneuralia.

The earliest history of onychophorans is central to this dilemma. The first scenario implies that foregut armature was present in the ancestral onychophoran, whereas under the second, onychophorans never had foregut armature. To reconstruct the ancestral configuration of the onychophoran foregut, we turned to the lobopodian *Hallucigenia sparsa*<sup>2–4</sup>, now regarded as a stem-group onychophoran<sup>5,6</sup>. Until now, this taxon's potential significance for early ecdysozoan evolution has been curtailed by uncertainty in its morphological interpretation: *Hallucigenia* has variously been reconstructed on its side, upside down and back to front (Extended Data Table 1). New material (Supplementary Table 1) and high-resolution microscopic analysis reveals many anatomical features in *Hallucigenia* for the first time. In particular, robust carbonaceous elements occur around *Hallucigenia*'s mouth and along its pharynx, implying that the ances-

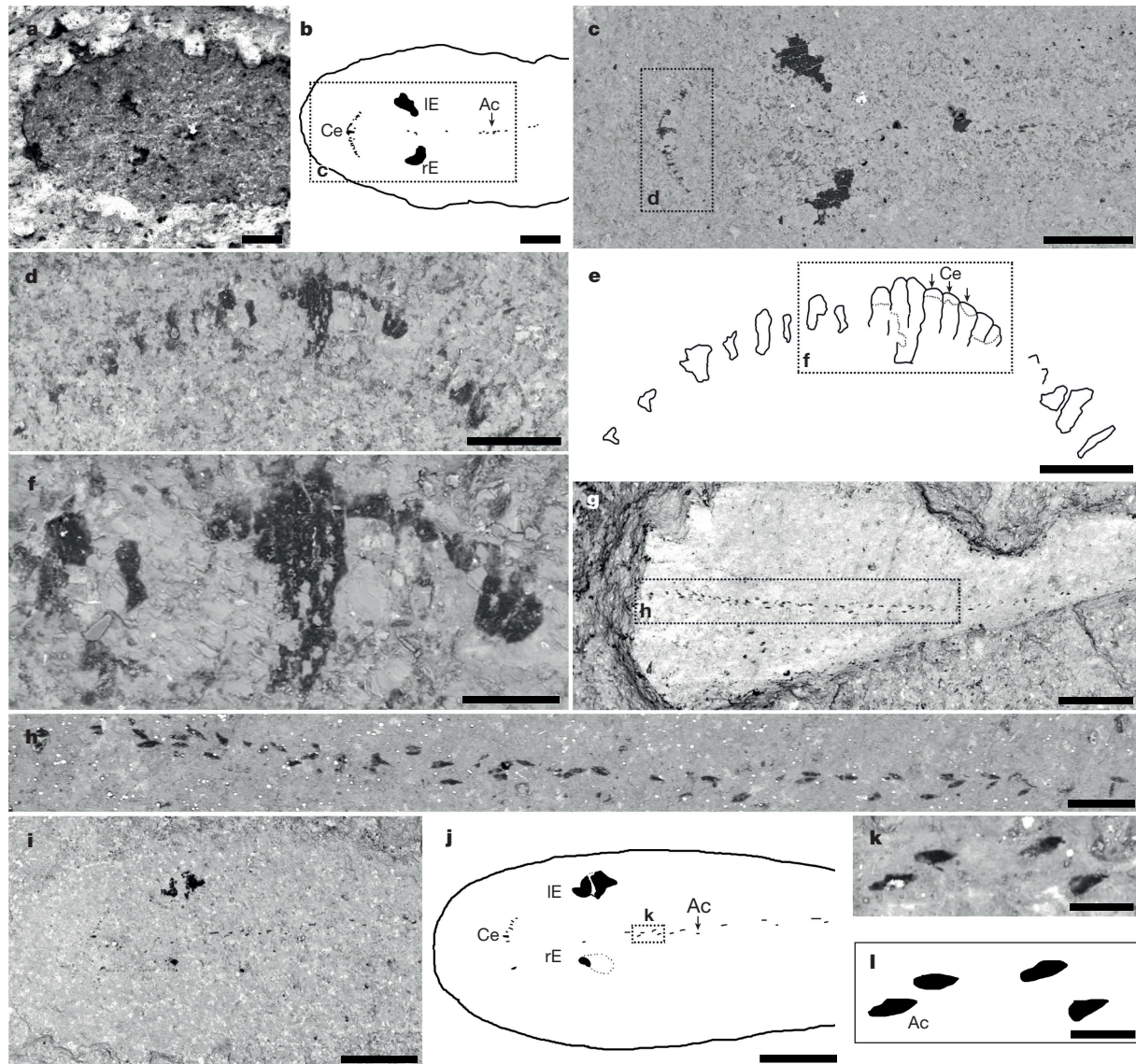


**Figure 1** | Optical images of *Hallucigenia sparsa* from the Burgess Shale (anterior to the left). **a**, ROM 62269 (see also Extended Data Fig. 5); **b**, ROM 63142; scanning electron microscope (SEM) images are provided in Fig. 2g–l and Extended Data Fig. 3c; **c**, ROM 63051; see also Extended Data Fig. 3b, d–g; **d**, ROM 63146; high-magnification images are provided in Fig. 2a–f and Extended Data Fig. 7; **e**, NMNH 198658; see also Extended Data Fig. 2b, c; **f**, **g**, anterior section of ROM 57168; see also Extended Data Fig. 1c–e. Acronyms for all figures: A, appendages; Ac, aciculae; An, anus; Bc, buccal

chamber; C, claw; Ce, circumoral elements; Cs, circumoral structure; Df, decay fluids; E, eyes; F, foregut; G, gut; L, left; Mo, mouth opening; r, right; S, spines; A<sub>1–n</sub> or S<sub>1–m</sub>, order of A or S from front to back. Dotted white lines identify areas enlarged in Fig. 2 and Extended Data figures, as denoted in captions. Unbroken white lines in **b–d** represent edges of the composite images of both parts and counterparts superimposed together. Black and white arrowhead denotes image flipped horizontally. Scale bars, 5 mm (**a–e**), 0.5 mm (**f**, **g**).

<sup>1</sup>Department of Earth Sciences, University of Cambridge, Cambridge CB2 3EQ, UK. <sup>2</sup>Department of Natural History (Palaeobiology Section), Royal Ontario Museum, Toronto, Ontario M5S 2C6, Canada. <sup>3</sup>Departments of Ecology and Evolutionary Biology and Earth Sciences, University of Toronto, Toronto, Ontario M5S 3B2, Canada.





**Figure 2 | Scanning electron micrographs of the head region of *Hallucigenia sparsa* from the Burgess Shale.** Anterior is to the left except for d–f, where anterior is to the top. a–f, ROM 63146 (see Fig. 1d) with sketches of anterior region (b) and circumoral elements (e); g–l, ROM 63142, part (g–h) and

counterpart (i–l), showing aciculae. See Fig. 1 for acronyms and symbols. Detector mode: a, secondary electron; c–k, backscatter. Scale bars, 200  $\mu$ m (a–c, g, i, j), 50  $\mu$ m (d–h), 20  $\mu$ m (f, k–l).

tral onychophoran—and seemingly the ancestral ecdysozoan—bore circumoral elements and pharyngeal teeth.

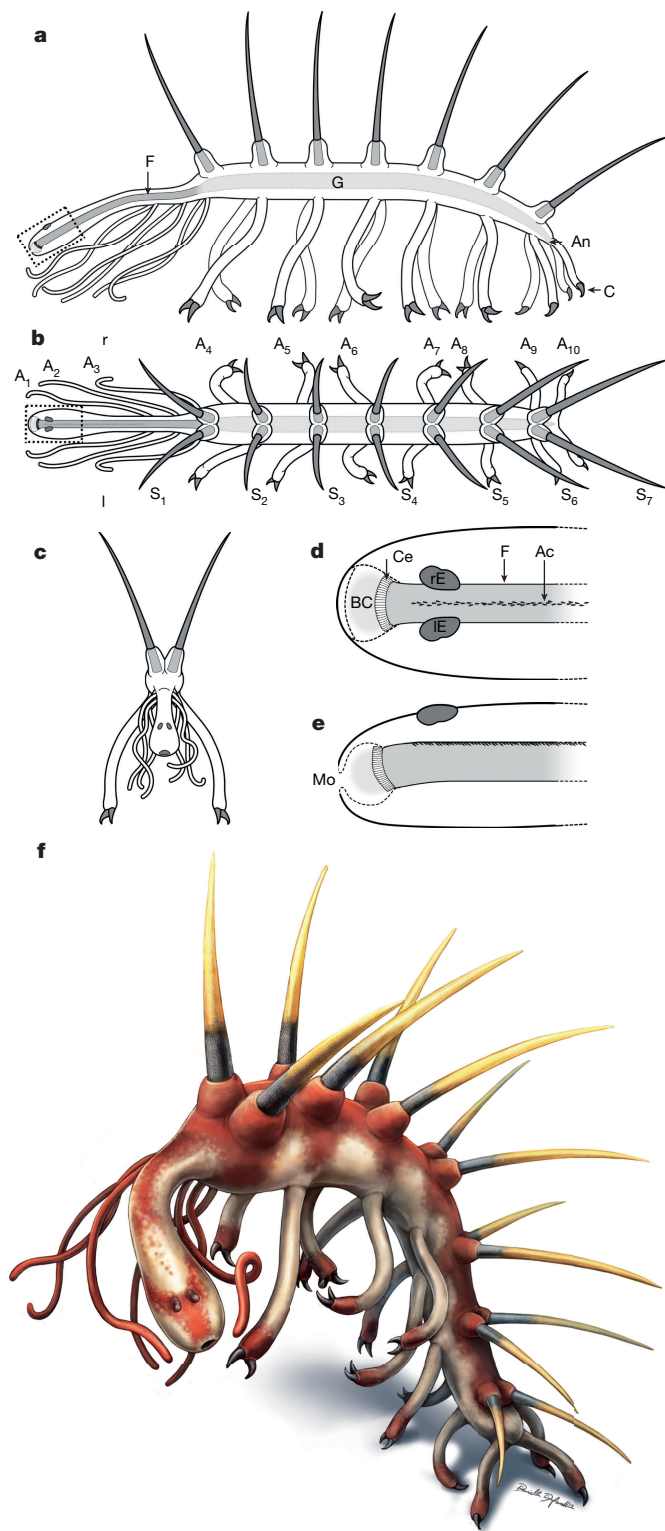
*Hallucigenia*'s tubular body ranges from 10 mm to more than 50 mm in length (Extended Data Fig. 1a–c and Supplementary Table 2). It bears ten elongate ventrolateral appendages (Fig. 1a–e); the anterior eight are of uniform length, whereas the posterior two are progressively shorter (Fig. 1d, e and Extended Data Fig. 2a–c). The final pair of appendages is terminal, confirming the absence of a posterior extension of the trunk<sup>4</sup>. The third to tenth appendage pairs are regularly spaced; the first, second and third appendage pairs are twice as close together (Fig. 1a, b, e and Extended Data Figs 1c, 3a, b, 4e and 5a). The anterior three pairs of appendages are 1.5–2.0 times narrower than the posterior seven, and lacked claws. These narrow appendages were flexible and long enough to reach the mouth (Fig. 1a, e–g and Extended Data Figs 1c, d, 2d, 3a, 4a and 6e, f). The posterior seven appendage pairs are legs with terminal claws: two claws are present on appendages four to eight, forming an acute angle (Fig. 1a–d and Extended Data Fig. 3c, d, g), whereas a single claw adorns appendages nine and ten.

Seven pairs of equally spaced elongate spines occupy the dorsolateral pinnacles of the trunk, situated above the third to ninth appendage

pairs (Fig. 1a–e). The spines in each pair are separated by 60–90° (Extended Data Figs 1, 4 and 7). Each spine is supported by a buttress of soft tissue which forms a hump-like swelling of the body wall and is particularly prominent in larger individuals (Fig. 1d and Extended Data Figs 1a, c, e and 6). The spines are uniform in length, width, spacing and shape: they are not quite straight but curve slightly ( $3.5^\circ \pm 0.9^\circ$ ) towards the posterior. The spines are centrifugally arranged in lateral view: the more anterior spines tilt forwards; the rear spines tilt backwards. The construction of the spines and claws from stacks of nested elements has been reported elsewhere<sup>5,6</sup>.

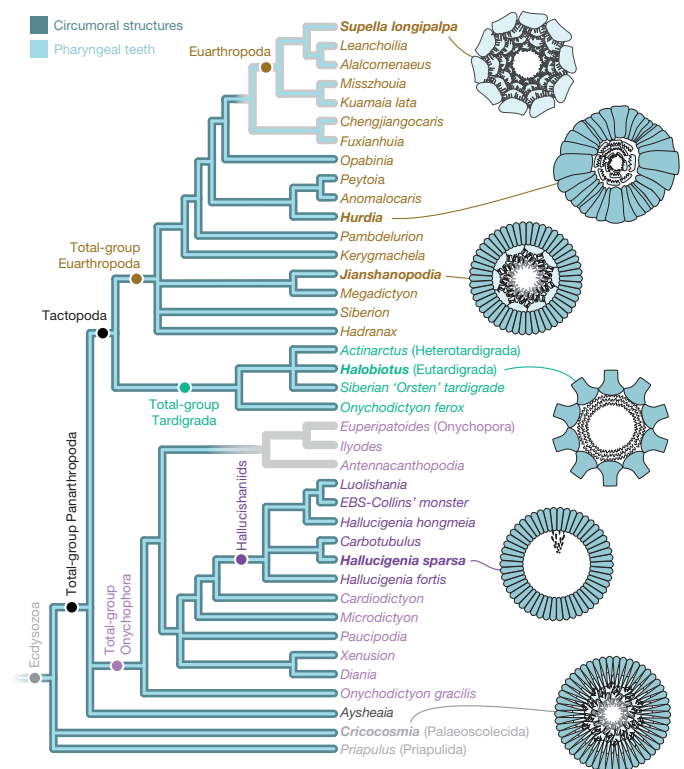
The character of the trunk changes markedly at the position of the first pair of spines. Behind this point, the trunk exhibits a uniform girth. (A linear relationship between trunk girth and body length indicates isometric growth; see Supplementary Table 2.) In front of the first spine pair, the trunk is a third narrower than the posterior trunk, with a bulbous anterior expansion evident in smaller specimens (Fig. 1a–e and Extended Data Figs 1–8). The anterior trunk usually bends at its midpoint, orienting the mouth opening ventrally.

Approximately 500  $\mu$ m from the anterior of the body and 100  $\mu$ m from the sagittal axis lies a dorsal pair of convex carbonaceous



**Figure 3 | Anatomical drawings of *Hallucigenia sparsa*.** **a**, Lateral profile; **b**, dorsal profile; **c**, frontal profile; **d**, **e**, head in dorsal (**d**) and lateral (**e**) views, corresponding to regions indicated in **b** and **a**, respectively; **f**, full reconstruction. Drawings, reproduced with permission, by Danielle Dufault. See Fig. 1 for acronyms.

impressions, reaching 200  $\mu\text{m}$  in diameter, which we interpret as eyes (Fig. 2a–c, i, j and Extended Data Figs 3, 5, 7 and 8b–d, i–m). Their irregular surface (Fig. 2c and Extended Data Figs 3e, 5f and 8j, m) argues against the presence of ommatidia; the eyes were presumably simple rather than compound. This seems to be consistent with the



**Figure 4 | Ecdysozoan phylogeny, showing most parsimonious character distribution of circumoral structures (dark blue) and pharyngeal teeth (light blue).** Fitch parsimony indicates the presence of both these structures in the ancestral ecdysozoan; a scenario positing multiple independent innovations of this armature would be less parsimonious. The topology shown denotes the strict consensus of all most parsimonious trees recovered under implied weights with concavity constant ( $k$ ) between 0.46 and 211, after the removal of *Orstenotubulus*. A newly retrieved 'hallucishaniid' clade—diagnosed by a swollen head, dorsal spines, and the differentiation of the anterior trunk and trunk appendages—includes luolishaniids, *Orstenotubulus* (not shown) and *Carbotubulus* within a paraphyletic 'Hallucigenia'. Illustrated taxa are in bold type; see discussion of transformation series 9 and 13 in Supplementary Note 1. For phylogenetic data and full results, see Supplementary Data.

eyes of other lobopodians (Supplementary Note 1, transformation series 18).

Reflective or darker regions occur along the axes of well-preserved appendages and appear, in the manner typical of lobopod limbs<sup>24</sup>, to represent extensions of the hydrostatic body cavity (Fig. 1e). A large ampulla-shaped structure that opens anteriorly represents a buccal chamber or 'mouth' (Fig. 1f, g and Extended Data Figs 1d, 2f, g, 4b, f and 8f, g), and is followed by a foregut that consistently occupies the central 50% of the anterior trunk (Fig. 1e and Extended Data Figs 1d, 2f, g, 4, 6, 7 and 8a, k). The foregut is darker than the surrounding tissue, conceivably indicating the presence of a cuticular lining. At the end of the head, the foregut widens into a broader, poorly preserved midgut (Fig. 1e and Extended Data Figs 2b, 4 and 6); the gut ends in a terminal anus (Extended Data Fig. 2b), through which decay fluids—represented by a darkly stained region of variable extent (Fig. 1b, e and Extended Data Figs 2a–c, 3a, b and 6a–d)—were expelled. Preservation of the hindgut is inadequate to determine whether it was differentiated from the midgut.

From behind the buccal chamber to the first pair of appendages, the dorsal surface of the foregut lumen is lined with dozens of posterior-directed aciculae (Fig. 2g–l and Extended Data Fig. 4c, d). These robustly carbonaceous structures are 10  $\mu\text{m}$  long and gently curved; their consistent size and orientation, uniform distribution, and absence elsewhere in the gut excludes the possibility that they represent gut contents; rather, they were biologically associated with the gut wall.



At the back of the buccal chamber, around 200 µm from the anterior termination of the trunk, lies a 250-µm-wide crescentic structure composed of multiple identical lamellae, each around 10 µm across and 60 µm long. Lamellae are evident in every structure that is preserved, and consistently display a radial arrangement (Fig. 2a–f, i, j and Extended Data Figs 5c, d and 8j–m). The structure is preserved laterally; it originally constituted a ring of lamellae around the opening of the foregut.

Like the claws and spines, the radial lamellae are preserved as discrete carbonaceous films—they were originally sclerotized, rather than representing soft tissue such as muscle, cuticular folds, or pigmentation, and they do not represent a taphonomic artefact. The lamellae are fundamentally unlike the modified pair of claws that form the jaws of modern onychophorans. Insofar as they are numerous, elongate, and sclerotized, and are arranged radially around the anterior opening of the foregut, the lamellae convincingly resemble the circumoral elements present in other ecdysozoans (see discussion in Supplementary Note 1, transformation series 9). To evaluate the evolutionary significance of this similarity we incorporated our observations (summarized in Fig. 3 and Supplementary Videos 1 and 2) into an updated phylogenetic matrix (Supplementary Data).

The reconstruction of character states through Fitch parsimony indicates that sclerotized circumoral elements were present in the ancestral ecdysozoan (Fig. 4 and Supplementary Note 1, transformation series 9), supporting the homology between circumoral structures in Tardigrada<sup>9,14</sup> and stem-euarthropods<sup>10,11,14,25</sup> and the circumoral ('coronal') spines of cycloneuralians<sup>13,20,26</sup> (see discussion in Supplementary Notes 1 and 2, transformation series 9). Homology between the panarthropod pharynx and the cycloneuralian pharynx is corroborated by the presence of robust sclerotized teeth in the anterior pharynx (Fig. 4 and Supplementary Note 1, transformation series 13), previously reported in extant cycloneuralians, euarthropods and tardigrades<sup>9,13,16,27</sup> and now also evident in stem-group onychophorans. The simple construction of the modern onychophoran foregut therefore reflects a secondary loss of cycloneuralian-like pharyngeal teeth and circumoral elements in the onychophoran stem lineage, and stands in marked contrast to the complex armoured foregut of the ancestral ecdysozoan.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 6 January; accepted 18 May 2015.

Published online 24 June 2015.

1. Aguinaldo, A. M. *et al.* Evidence for a clade of nematodes, arthropods and other moulting animals. *Nature* **387**, 489–493 (1997).
2. Conway Morris, S. A new metazoan from the Cambrian Burgess Shale of British Columbia. *Palaeontology* **20**, 623–640 (1977).
3. Ramsköld, L. & Chen, J.-Y. in *Arthropod Fossils and Phylogeny* (ed. Edgecombe, G. D.) 107–150 (Columbia Univ. Press, 1998).
4. Ramsköld, L. The second leg row of *Hallucigenia* discovered. *Lethaia* **25**, 221–224 (1992).
5. Caron, J.-B., Smith, M. R. & Harvey, T. H. P. Beyond the Burgess Shale: Cambrian microfossils track the rise and fall of hallucigeniid lobopodians. *Proc. R. Soc. Lond. B* **280**, 20131613 (2013).
6. Smith, M. R. & Ortega-Hernández, J. *Hallucigenia*'s onychophoran-like claws and the case for Tactopoda. *Nature* **514**, 363–366 (2014).

7. Dewel, R. A. & Clark, W. H. Studies on the tardigrades. II. Fine structure of the pharynx of *Milnesium tardigradum* Doyère. *Tissue Cell* **5**, 147–159 (1973).
8. Dewel, R. A. & Eibye-Jacobsen, J. The mouth cone and mouth ring of *Echiniscus viridissimus* Peterfi, 1956 (Heterotardigrada) with comparisons to corresponding structures in other tardigrades. *Hydrobiologia* **558**, 41–51 (2006).
9. Guidetti, R. *et al.* Form and function of the feeding apparatus in Eutardigrada (Tardigrada). *Zoomorphology* **131**, 127–148 (2012).
10. Vannier, J., Liu, J., Lerosey-Aubril, R., Vinther, J. & Daley, A. C. Sophisticated digestive systems in early arthropods. *Nat. Commun.* **5**, 3641 (2014).
11. Budd, G. E. The morphology and phylogenetic significance of *Kerygmachela kierkegaardii* Budd (Buen Formation, Lower Cambrian, N Greenland). *Trans. R. Soc. Edinb. Earth Sci.* **89**, 249–290 (1998).
12. Daley, A. C., Budd, G. E., Caron, J.-B., Edgecombe, G. D. & Collins, D. H. The Burgess Shale anomalocaridid *Hurdia* and its significance for early euarthropod evolution. *Science* **323**, 1597–1600 (2009).
13. Conway Morris, S. Fossil priapulid worms. *Spec. Pap. Pal.* **20** (1977).
14. Budd, G. E. Tardigrades as 'stem-group arthropods': the evidence from the Cambrian fauna. *Zool. Anz.* **240**, 265–279 (2001).
15. Eriksson, B. J. & Budd, G. E. Onychophoran cephalic nerves and their bearing on our understanding of head segmentation and stem-group evolution of Arthropoda. *Arthropod Struct. Dev.* **29**, 197–209 (2000).
16. Elzinga, R. J. Microspines in the alimentary canal of Arthropoda, Onychophora, Annelida. *Int. J. Ins. Morph. Emb.* **27**, 341–349 (1998).
17. Dewel, R. A. & Dewel, W. C. The brain of *Echiniscus viridissimus* Peterfi, 1956 (Heterotardigrada): A key to understanding the phylogenetic position of tardigrades and the evolution of the arthropod head. *Zool. J. Linn. Soc.* **116**, 35–49 (1996).
18. Waggoner, B. The earliest evolutionary history of arthropods: what can the fossils tell us? *Bol. Soc. Entomológica Aragon.* **26**, 115–131 (1999).
19. Hou, X.-G. *et al.* *Anomalocaris* and other large animals in the lower Cambrian Chengjiang fauna of southwest China. *GFF* **117**, 163–183 (1995).
20. Conway Morris, S. & Peel, J. S. New palaeoscolecoidan worms from the Lower Cambrian: Sirius Passet, Latham Shale, and Kinzers Shale. *Acta Pal. Pol.* **55**, 141–156 (2010).
21. Rota-Stabelli, O. *et al.* A congruent solution to arthropod phylogeny: phylogenomics, microRNAs and morphology support monophyletic Mandibulata. *Proc. R. Soc. Lond. B* **278**, 298–306 (2011).
22. Martin, C. & Mayer, G. Neuronal tracing of oral nerves in a velvet worm—implications for the evolution of the ecdysozoan brain. *Front. Neuroanat.* **8**, 7 (2014).
23. Dunn, C. W., Giribet, G., Edgecombe, G. D. & Hejnol, A. Animal phylogeny and its evolutionary implications. *Annu. Rev. Ecol. Syst.* **45**, 371–395 (2014).
24. Liu, J. *et al.* New observations of the lobopod-like worm *Facivermis* from the Early Cambrian Chengjiang Lagerstätte. *Chin. Sci. Bull.* **51**, 358–363 (2006).
25. Daley, A. C. & Bergström, J. The oral cone of *Anomalocaris* is not a classic 'peytoia'. *Naturwissenschaften* **99**, 501–504 (2012).
26. Maas, A., Huang, D., Chen, J.-Y., Waloszek, D. & Braun, A. Maotianshan-Shale nematelmiths—morphology, biology, and the phylogeny of Nematelmithes. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **254**, 288–306 (2007).
27. Kristensen, R. M. Loricifera, a new phylum with Aschelminthes characters from the meiobenthos. *Zeit. Zool. Syst. Evol.* **21**, 163–180 (1983).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank D. Erwin, M. Florence and P. Fenton for access to material and assistance with collections, G. Kretschmann and S. Whittaker for assistance with electron microscopy, and K. Meechan for assistance with data collection. ROM specimens were collected under Parks Canada Research and Collection permits to D. Collins. TNT is made available with the sponsorship of the Willi Hennig Society. Funding for this research was provided by Clare College, Cambridge (M.R.S.), NSERC Discovery Grant no. 341944 (J.-B.C.) and the ROM DMV Acquisition and Research Fund and Endowment Fund (J.-B.C.). This is ROM Burgess Shale project number 60.

**Author Contributions** M.R.S. performed data analysis; both authors examined and imaged material and contributed to the writing of the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.R.S. (ms609@cam.ac.uk).



## METHODS

**Fossil materials.** Materials are deposited at the Royal Ontario Museum, Toronto (ROM) and the Smithsonian Institution National Museum of Natural History, Washington DC (NMNH). Sediment covering parts of certain ROM specimens was mechanically removed using a tungsten-tipped micro-engraving tool. Specimens were photographed under various lighting conditions including dark- and bright-field illumination and polarized light, and imaged by backscatter and secondary electron microscopy under variable pressure.

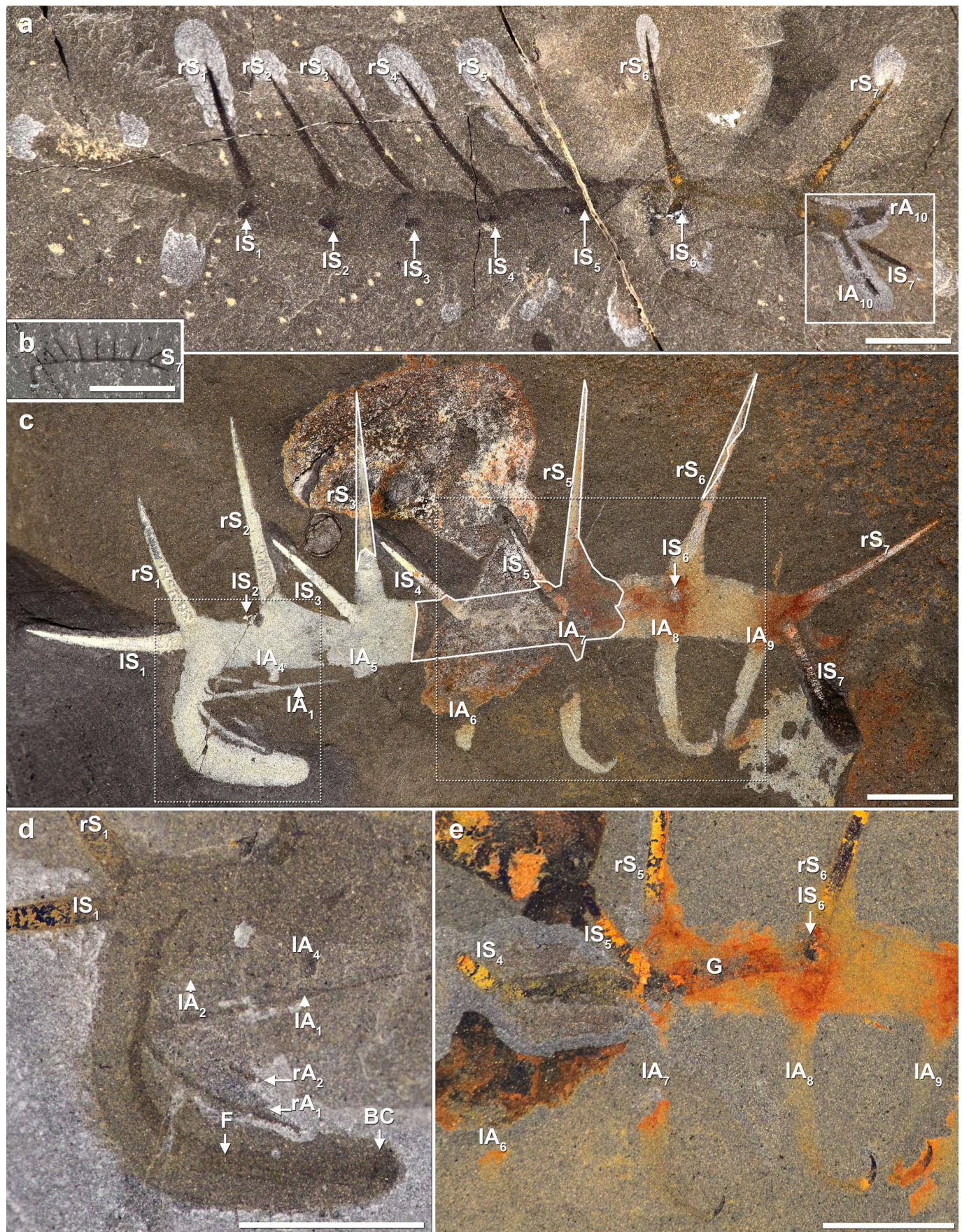
**Taphonomic considerations.** As with other Burgess Shale organisms<sup>28,29</sup>, *Hallucigenia sparsa* exhibits various degrees of pre- and post-burial decay, ranging from disarticulated specimens represented only by pairs of decay-resistant spines (Extended Data Fig. 9a), through partly disarticulated specimens retaining parts of the body (Extended Data Fig. 9b), to complete specimens whose curled appendages and trunks are consistent with post-mortem contraction following rapid burial of live organisms (Fig. 1a–e and Extended Data Figs 1–8). Consequently, the widths of the trunk and appendages are subject to slight taphonomic variation within and between specimens (as in, for example, Fig. 1). The full length of the body and appendages, where preserved, is typically buried within the matrix and is difficult to prepare mechanically.

**Phylogenetic analysis.** Phylogenetic analysis was conducted using the methods of Smith and Ortega-Hernández<sup>6</sup>; in summary, parsimony analysis was performed in

TNT<sup>30</sup> under a range of weighting parameters, with Goloboff's concavity constant<sup>31</sup> ranging from  $k = 0.118$  to 211, and under equal weights ( $k = \infty$ ). Code is available in the Supplementary Data. *Orstenotubulus* (80% tokens 'ambiguous' or 'inapplicable') was identified as a wildcard taxon with an unconstrained position within the hallucigeniids; to improve resolution it is omitted from the strict consensus trees presented in the main manuscript.

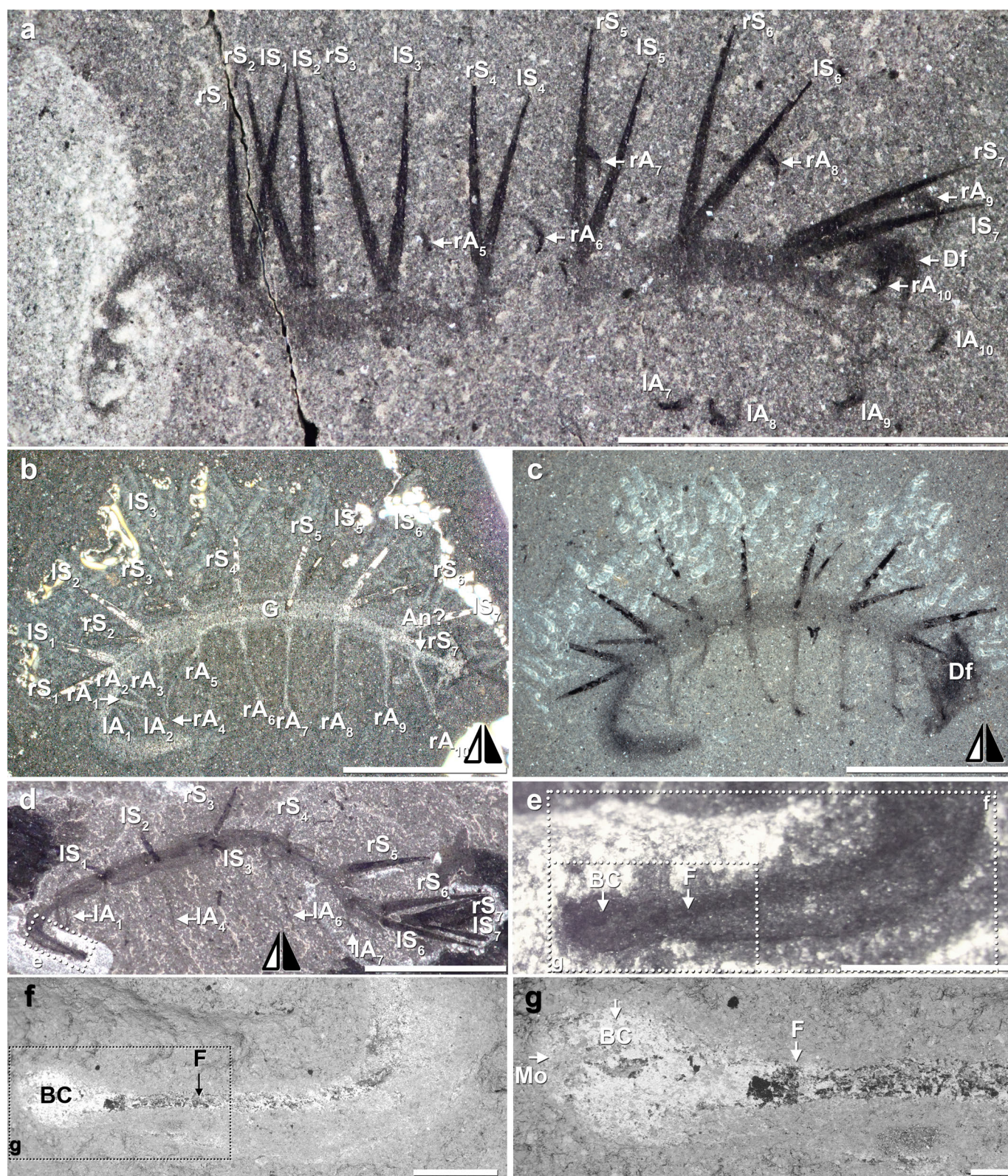
28. Conway Morris, S. & Caron, J.-B. *Pikaia gracilens* Walcott, a stem-group chordate from the Middle Cambrian of British Columbia. *Biol. Rev. Camb. Philos. Soc.* **87**, 480–512 (2012).
29. Smith, M. R. Nectocaridid ecology, diversity and affinity: early origin of a cephalopod-like body plan. *Paleobiology* **39**, 297–321 (2013).
30. Goloboff, P. A., Farris, J. S. & Nixon, K. C. TNT, a free program for phylogenetic analysis. *Cladistics* **24**, 774–786 (2008).
31. Goloboff, P. A. Estimating character weights during tree search. *Cladistics* **9**, 83–91 (1993).
32. Walcott, C. D. Cambrian Geology and Paleontology II, no. 5. Middle Cambrian annelids. *Smithson. Misc. Collect.* **57**, 109–144 (1911).
33. Ramsköld, L. & Hou, X.-G. New early Cambrian animal and onychophoran affinities of enigmatic metazoans. *Nature* **351**, 225–228 (1991).
34. Hou, X.-G. & Bergström, J. Cambrian lobopodians—ancestors of extant onychophorans? *Zool. J. Linn. Soc.* **114**, 3–19 (1995).
35. Steiner, M., Hu, S.-X., Liu, J. & Keupp, H. A new species of *Hallucigenia* from the Cambrian Stage 4 Wulongqing Formation of Yunnan (South China) and the structure of sclerites in lobopodians. *Bull. Geosci.* **87**, 107–124 (2012).





**Extended Data Figure 1** | *Hallucigenia sparsa* from the Burgess Shale. **a, b**, largest (**a**, ROM 57169) and smallest (**b**, ROM 62093) specimens, to the same scale; **c**, ROM 57168, with enlargements of the anterior (**d**) and mid-trunk (**e**). Acronyms as in Fig. 1. Scale bars, 5 mm.

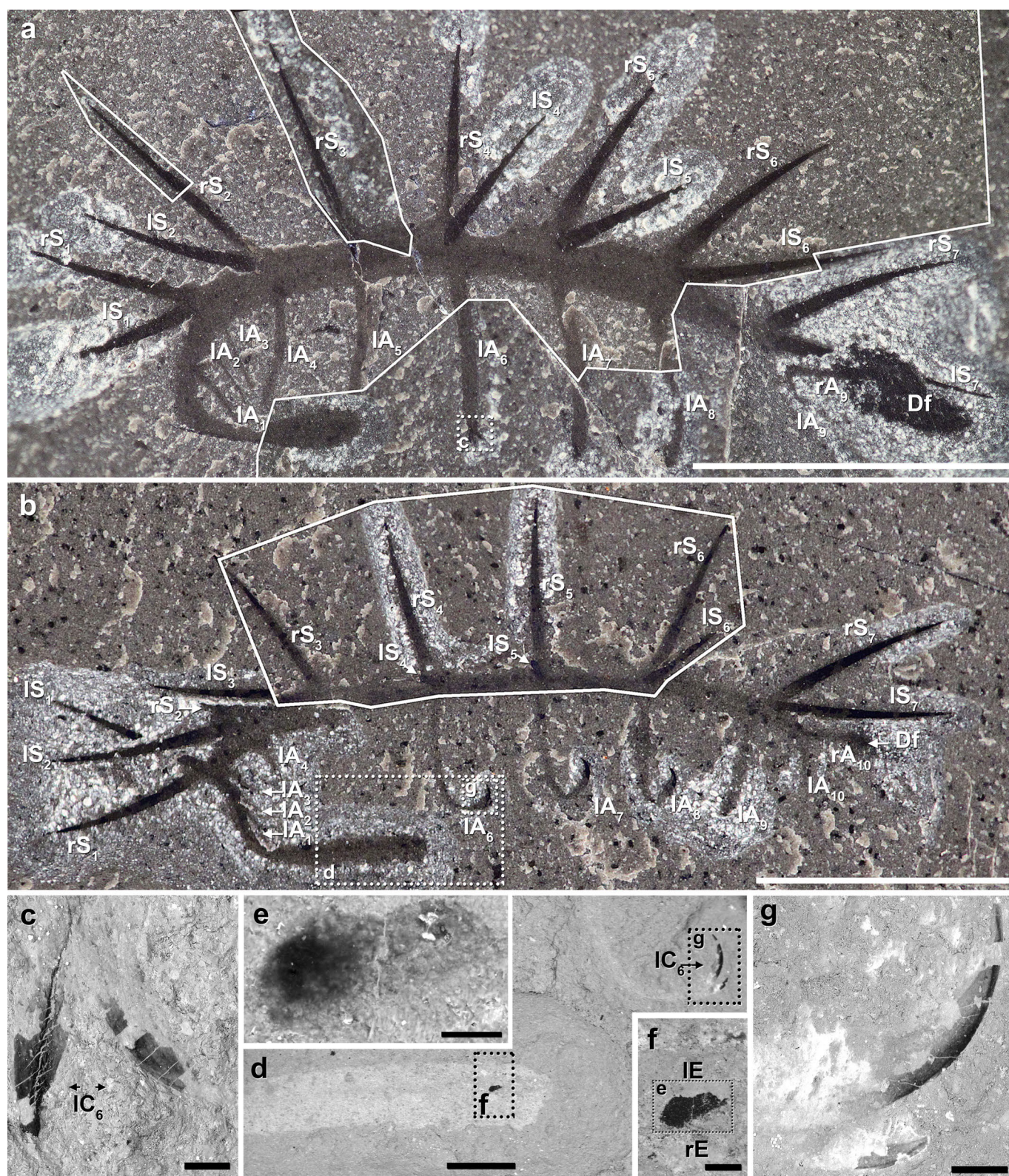




**Extended Data Figure 2 | *Hallucigenia sparsa* from the Burgess Shale.**  
 a, ROM 63139, showing posterior body termination; b, c, NMNH 198658, showing posterior termination (see also Fig. 1e); d–g, ROM 63143:

e, enlargement of region marked in d; f, g, backscatter SEMs of regions marked in e. Acronyms as in Fig. 1. Scale bars, 5 mm (a–d), 1 mm (e), 0.5 mm (f), 0.1 mm (g).

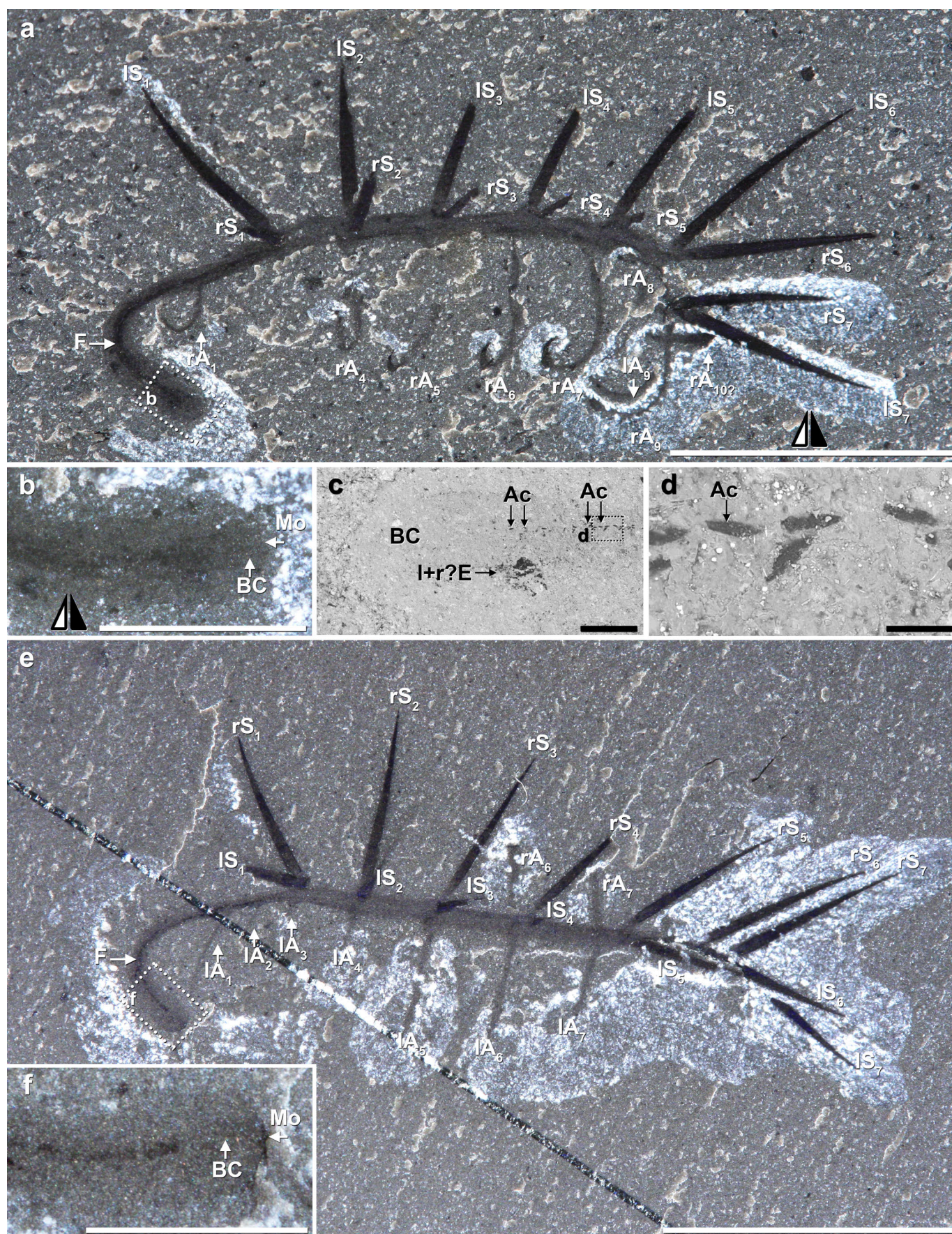




**Extended Data Figure 3 | *Hallucigenia sparsa* from the Burgess Shale.**  
**a, c, ROM 63142:** **a**, composite image incorporating part and counterpart of the entire specimen; **c**, claw pair. **b, d–g, ROM 63051:** **b**, composite image

incorporating part and counterpart of the entire specimen; **d**, anterior section; **e, f**, eyes; **g**, claw pair. **c–e** are backscatter electron micrographs. Acronyms as in Fig. 1. Scale bars, 5 mm (**a, b**), 500  $\mu$ m (**d**), 50  $\mu$ m (**c, f, g**), 20  $\mu$ m (**e**).



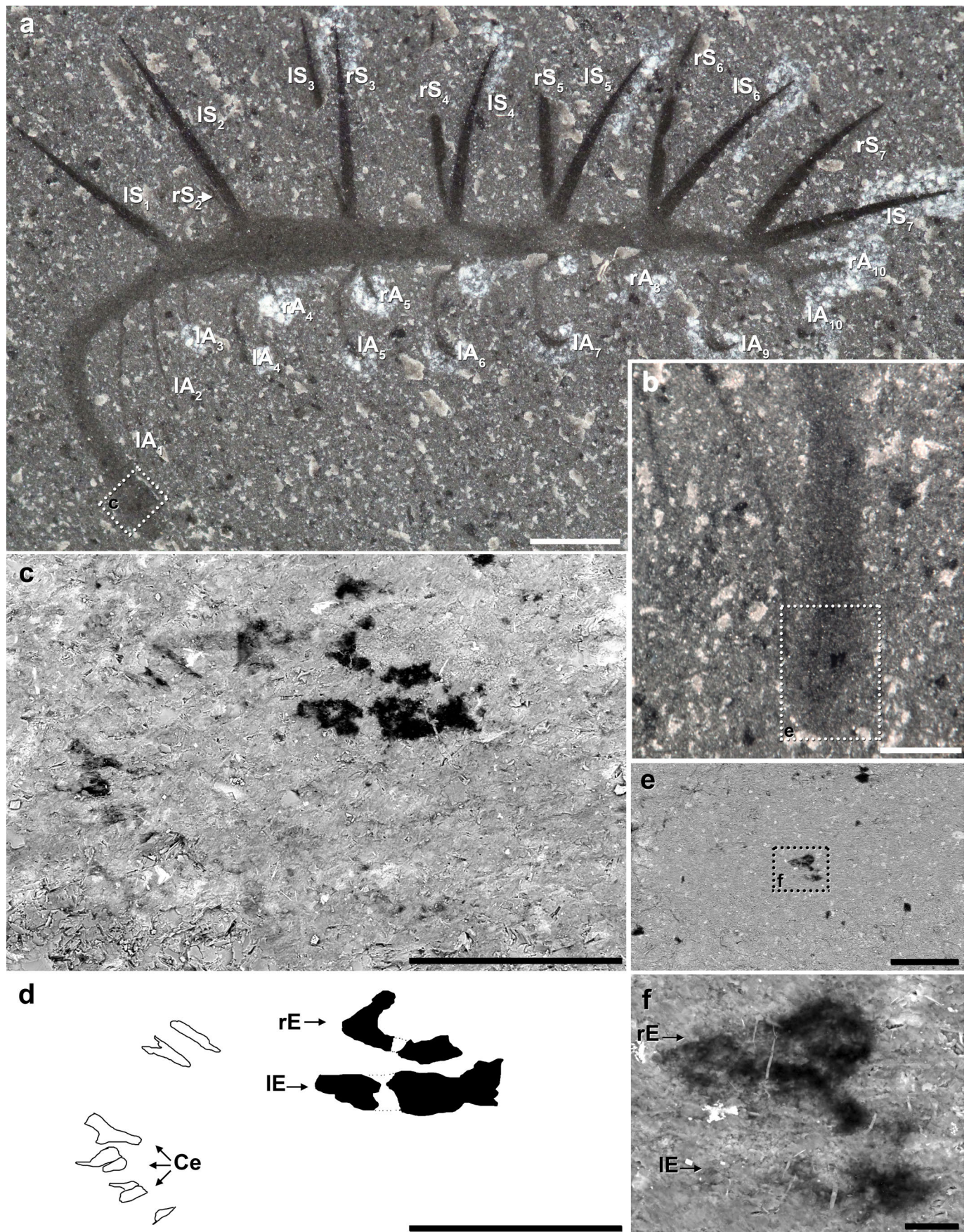


**Extended Data Figure 4 | *Hallucigenia sparsa* from the Burgess Shale.**

**a–d**, ROM 61513: **a**, entire specimen; **b–d**, enlargements of anterior region, showing mouth opening, aciculae and eyes; mouth opening to right in **b**, to left

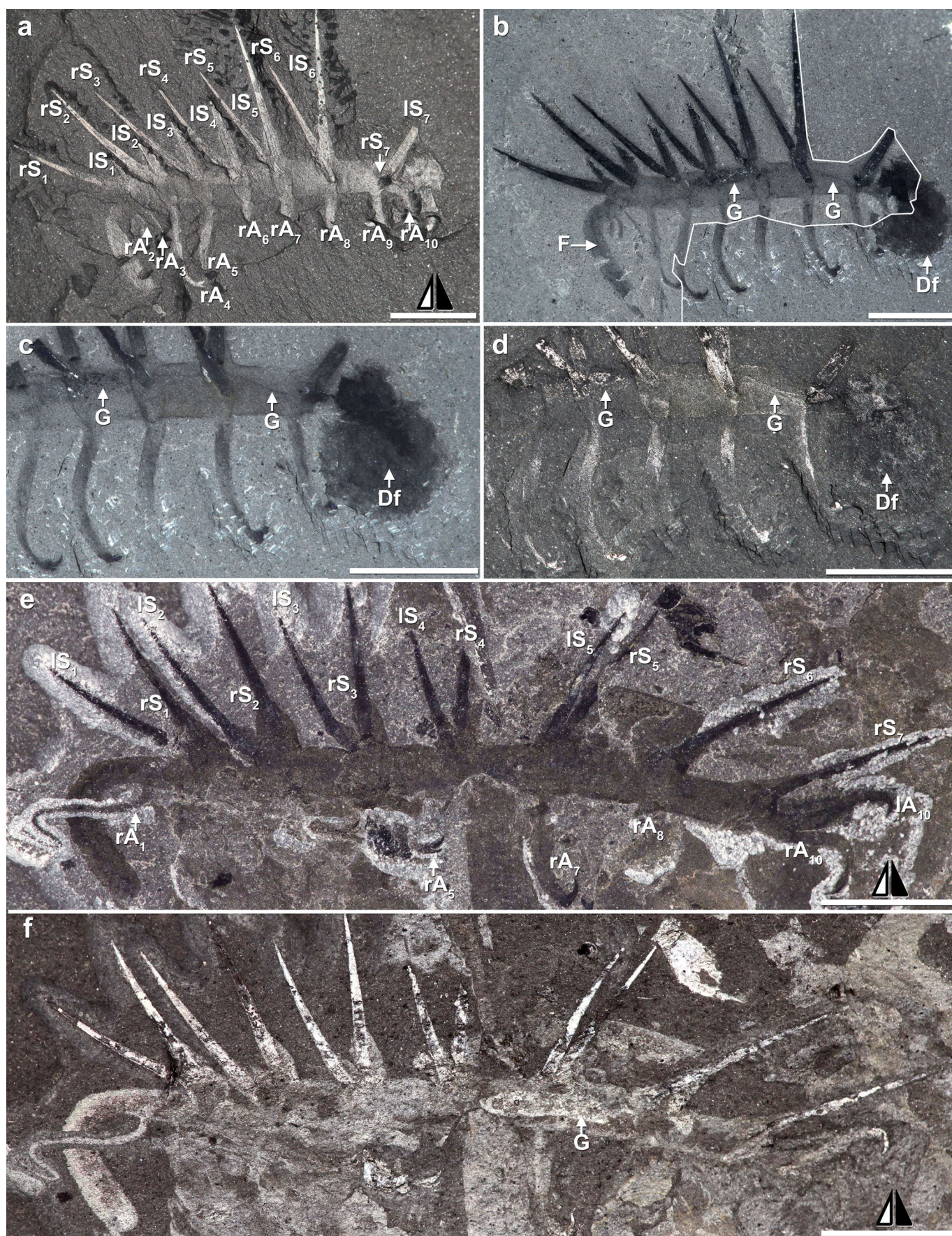
in **c**, **d**, **e**, **f**, ROM 61143: anterior region marked in **e** is enlarged in **f**. Acronyms as in Fig. 1. Scale bars, 5 mm (**a**, **e**), 1 mm (**b**, **f**), 200  $\mu$ m (**c**), 20  $\mu$ m (**d**).





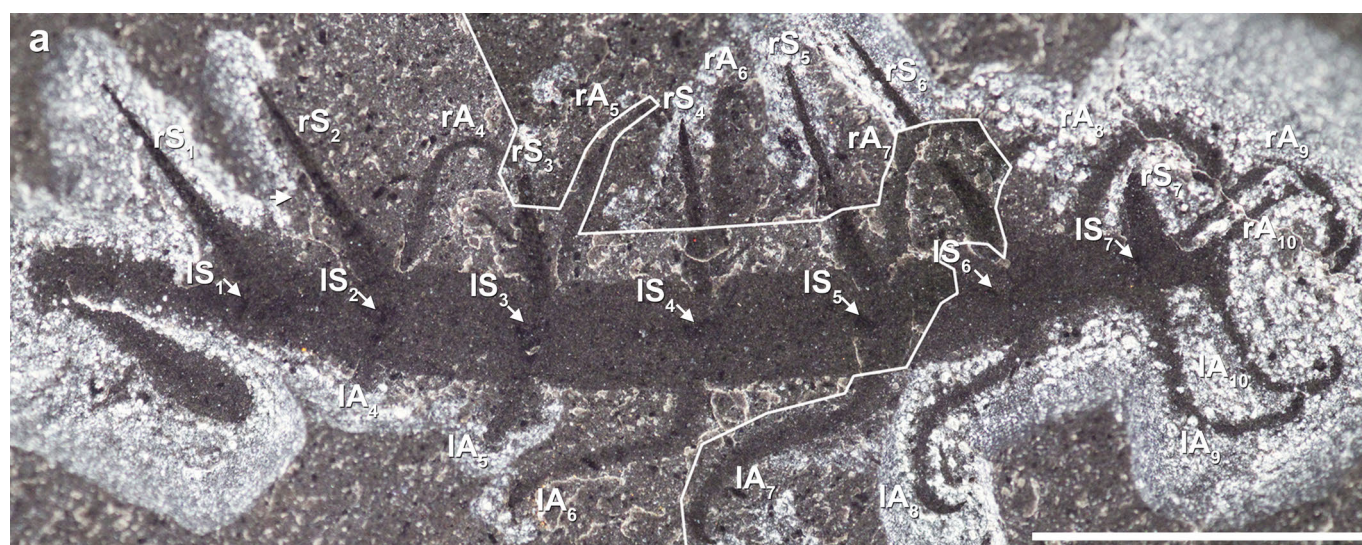
**Extended Data Figure 5** | *Hallucigenia sparsa* (ROM 62269) from the Burgess Shale. **a**, part; **b**, counterpart, anterior section, showing eyes; **c**, **d**, eyes and mouthparts (backscatter SEM); **e**, **f**, detail of eyes (counterpart). Acronyms as in Fig. 1. Scale bars, 1 mm (**a**), 400  $\mu$ m (**b**), 200  $\mu$ m (**e**), 100  $\mu$ m (**c**, **d**), 20  $\mu$ m (**f**).





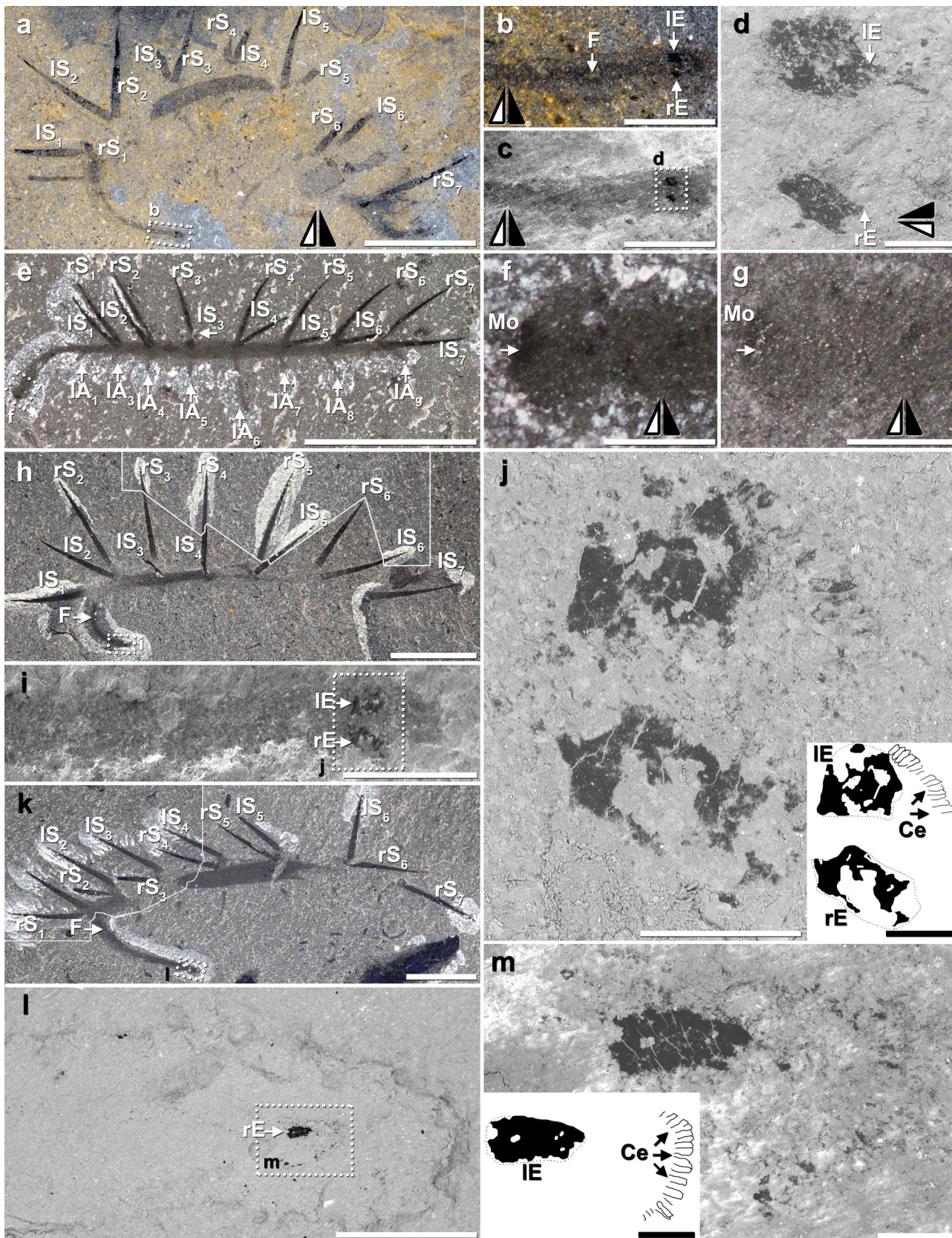
**Extended Data Figure 6** | *Hallucigenia sparsa* from the Burgess Shale. **a–d**, NMNH 83935 (holotype): in contrast to body tissue, decay fluids lack a sharp margin and are non-reflective; **e, f**, ROM 57776, showing full length of appendage one. Acronyms as in Fig. 1. Scale bars, 5 mm.





Extended Data Figure 7 | *Hallucigenia sparsa* (ROM 63146), composite image of part and counterpart. Acronyms as in Fig. 1. Scale bar, 5 mm.



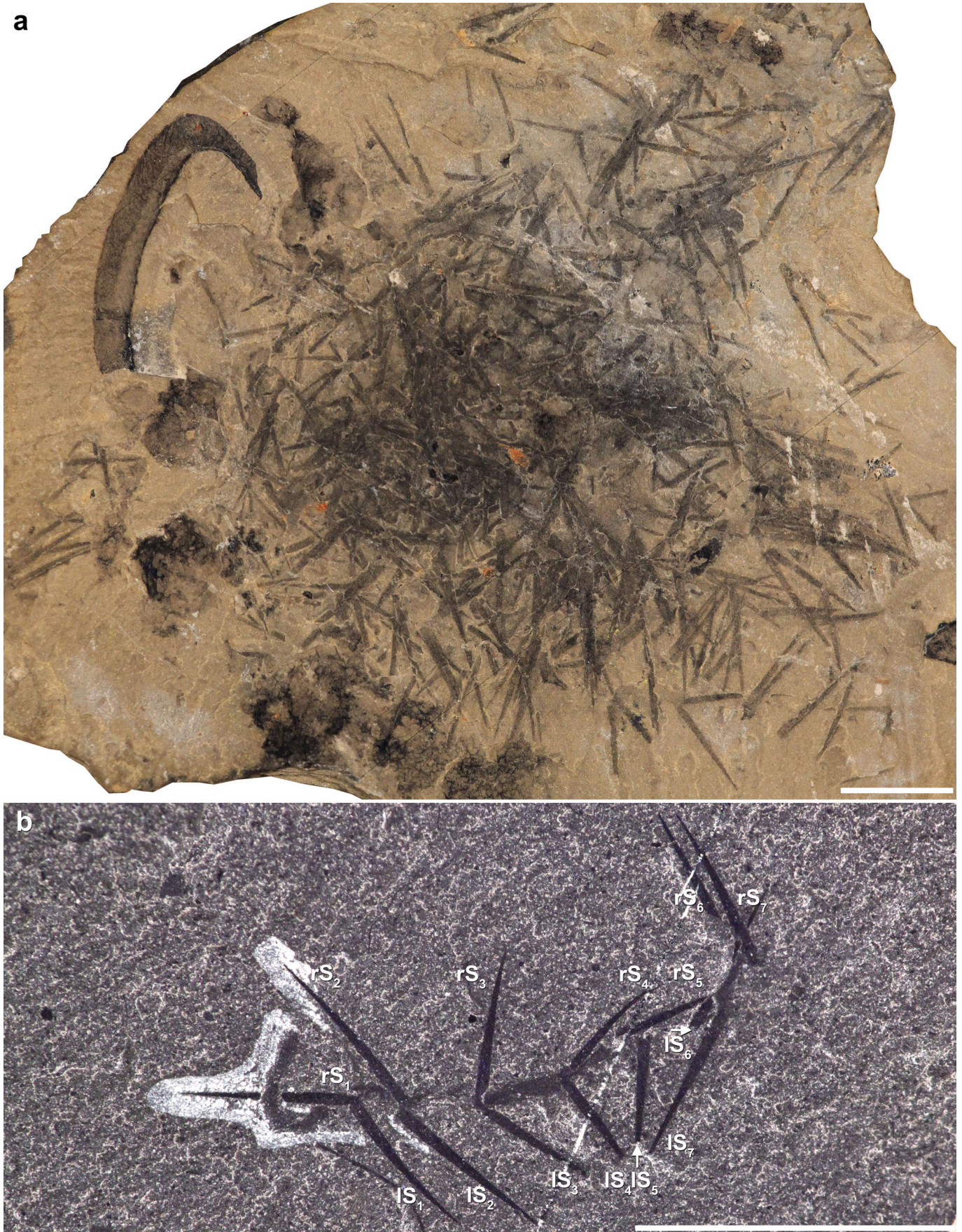
Extended Data Figure 8 | *Hallucigenia sparsa* from the Burgess Shale.

**a–d**, NMNH 193996: **b**, **c**, enlargements of area boxed in **a**; **c**, secondary electron micrograph; **d**, backscatter electron micrograph of region marked in **c**. **e–g**, ROM 63141, showing position of mouth. **h–j**, ROM 63144: **i**, secondary electron image of region marked in **h**; **j**, backscatter electron image of region

marked in **i**, showing eyes and mouthparts, with interpretative diagram.

**k–m**, ROM 63140: **l**, backscatter SEM of head, showing right eye and mouthparts (enlarged in **m**, with interpretative diagram). Acronyms as in Fig. 1. Scale bars, 10 mm (**k**), 5 mm (**a**, **e**, **h**), 1 mm (**b**, **c**, **l**), 0.5 mm (**i**), 0.1 mm (**d**, **j**, **m**).





**Extended Data Figure 9 | *Hallucigenia sparsa* from the Burgess Shale.** **a**, ROM 43045, cluster of dissociated specimens; **b**, ROM 63145, dissociated specimen showing spines in close anatomical position. Acronyms as in Fig. 1. Scale bars, 10 mm.



Extended Data Table 1 | Interpretations of *Hallucigenia* through time

| Authority:  | Walcott 1911 <sup>32</sup> | Conway Morris 1977 <sup>2</sup> | Ramsköld & Hou 1991 <sup>33</sup> | Ramsköld 1992 <sup>4</sup> | Hou & Bergström 1995 <sup>34</sup> | Steiner <i>et al.</i> 2012 <sup>35</sup> | This study                        |
|---|----------------------------|---------------------------------|-----------------------------------|----------------------------|------------------------------------|--|-----------------------------------|
| <b>Spines:</b>  | Lateral (parapodia)        | Ventral                         | Dorsal                            | Dorsal                     | Dorsal                             | Dorsal                                   | Dorsal                            |
| <b>Trunk appendages:</b>  | Not interpreted            | Single                          | Paired?                           | Paired                     | Paired                             | Paired                                   | Paired                            |
| <b>Slender legs at:</b>   | Not observed               | Posterior                       | Posterior                         | Anterior                   | Posterior                          | Anterior                                 | Anterior                          |
| <b>Pairs of slender legs:</b>                                   | Not observed               | 3                               | 3                                 | 3                          | 1 pair in <i>H. fortis</i>         | 2 (in <i>H. hongmeia</i> )               | 3                                 |
| <b>Width of trunk beyond slender legs:</b>                      | Not observed               | Constant                        | Narrowing slightly                | Narrowing slightly         | Bulge                              | Bulge                                    | Narrow 'neck' with terminal bulge |
| <b>Vertical trunk curvature beyond slender legs:</b>            | Not observed               | Curved                          | Straight                          | Straight                   | Straight                           | Straight                                 | Curved ventrally                  |
| <b>Position of first slender leg:</b>                           | Not observed               | Beyond spine                    | Beyond spine                      | Below end spine            | Beyond spine                       | Beyond spine                             | Below first spine                 |
| <b>(Pairs of) legs beyond spines at opposite end:</b>           | Not observed               | One [pair]                      | One pair                          | One pair                   | One pair                           | One pair                                 | One pair                          |
| <b>Body continues beyond legs at end opposite slender legs:</b> | Not observed               | Yes                             | Yes                               | No                         | Yes                                | Yes                                      | No                                |

References 32–35 are listed in the Methods section.

# Sex reversal triggers the rapid transition from genetic to temperature-dependent sex

Clare E. Holleley<sup>1</sup>, Denis O'Meally<sup>1</sup>, Stephen D. Sarre<sup>1</sup>, Jennifer A. Marshall Graves<sup>1,2</sup>, Tariq Ezaz<sup>1</sup>, Kazumi Matsubara<sup>1†</sup>, Bhumika Azad<sup>1†</sup>, Xiuwen Zhang<sup>1</sup> & Arthur Georges<sup>1</sup>

Sex determination in animals is amazingly plastic. Vertebrates display contrasting strategies ranging from complete genetic control of sex (genotypic sex determination) to environmentally determined sex (for example, temperature-dependent sex determination)<sup>1</sup>. Phylogenetic analyses suggest frequent evolutionary transitions between genotypic and temperature-dependent sex determination in environmentally sensitive lineages, including reptiles<sup>2</sup>. These transitions are thought to involve a genotypic system becoming sensitive to temperature, with sex determined by gene–environment interactions<sup>3</sup>. Most mechanistic models of transitions invoke a role for sex reversal<sup>3–5</sup>. Sex reversal has not yet been demonstrated in nature for any amniote, although it occurs in fish<sup>6</sup> and rarely in amphibians<sup>7,8</sup>. Here we make the first report of reptile sex reversal in the wild, in the Australian bearded dragon (*Pogona vitticeps*), and use sex-reversed animals to experimentally induce a rapid transition from genotypic to temperature-dependent sex determination. Controlled mating of normal males to sex-reversed females produces viable and fertile offspring whose phenotypic sex is determined solely by temperature (temperature-dependent sex determination). The W sex chromosome is eliminated from this lineage in the first generation. The instantaneous creation of a lineage of ZZ temperature-sensitive animals reveals a novel, climate-induced pathway for the rapid transition between genetic and temperature-dependent sex determination, and adds to concern about adaptation to rapid global climate change.

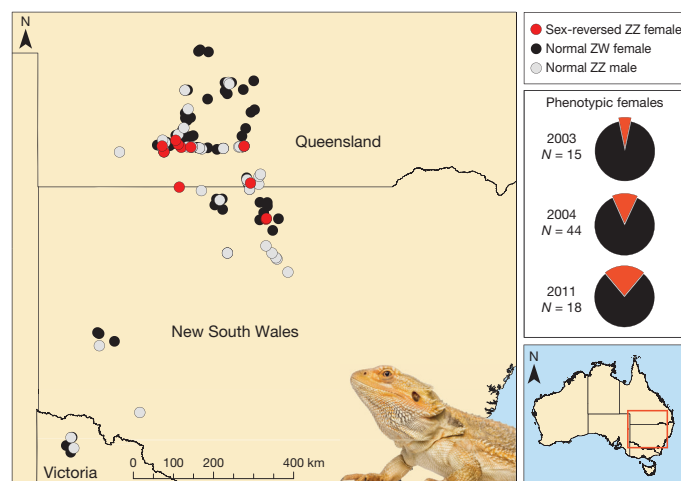
Sex determination is the regulatory process that initiates differentiation of the gonads in the early embryo to form either testes or ovaries. Very many reptiles have temperature-dependent sex determination (TSD), whereby the temperature that eggs experience in the nest determines the sex of offspring. Others have male or female heterogamety, either XX/XY systems as in mammals or ZZ/ZW systems as in birds, with or without strongly heteromorphic sex chromosomes<sup>9</sup>. Mounting evidence suggests that genotypic and environmental modes of sex determination are not mutually exclusive dichotomous strategies<sup>1</sup>. Many species have differentiated sex chromosomes, but also show a temperature override, where genes and environment interact to determine sex<sup>10–14</sup>. Furthermore, the great diversity of sex-determining mechanisms seen in reptiles, matched in amphibians and fishes, but not mammals and birds, shows poor respect for phylogeny, implying a complex evolutionary history of multiple transitions among sex determination modes<sup>2,9</sup>.

The widely distributed Australian central bearded dragon (*P. vitticeps*) was one of the first reptile species in which a temperature override was observed<sup>10–12,14</sup>, and the first in which it was demonstrated genetically<sup>12</sup>. *P. vitticeps* has a female heterogametic system of sex determination with ZZ males (ZZm) and ZW females (ZWf)<sup>15</sup>, but high incubation temperatures experimentally feminize chromosomally male animals and produce sex-reversed females (ZZf)<sup>12</sup>. To identify sex-reversed females, we developed a new robust sex-specific

molecular marker from previously characterized *P. vitticeps* sex chromosome sequences<sup>16,17</sup> and validated the test against a panel of unrelated individuals incubated at temperatures where phenotypic sex and genotypic sex are concordant (20 ZZm and 20 ZWf). Absence of a W chromosome in putatively sex-reversed ZZ females was confirmed cytogenetically to eliminate the possibility of low-frequency recombination being mistakenly interpreted as sex reversal (Extended Data Figs 1–3).

Animals were sampled from several widely distributed populations. Application of the PCR sex marker to 131 wild-caught individuals identified 11 sex-reversed ZZ females occurring towards the northern end of the *P. vitticeps* range, near the border of Queensland and New South Wales (Fig. 1). Although sex reversal in reptiles has been demonstrated under laboratory conditions<sup>10–12</sup>, this is the first time that sex reversal has been shown to occur naturally in a wild population of reptiles, or indeed any amniote. Sex reversal was widespread in this population, with instances distributed over a total area of 23,650 km<sup>2</sup> in remote semi-arid Australia. Among wild phenotypic females, the proportion of ZZ sex-reversed females increased each year over the study, from 6.7% in 2003, to 13.6% in 2004, to 22.2% in 2011 (Fig. 1), suggestive of a trend but not significant ( $\chi^2 = 1.65$ , d.f. = 2,  $P = 0.44$ ).

The sex-reversed females were viable and fertile. In fact, our sex-reversed females laid significantly more eggs per year (mean

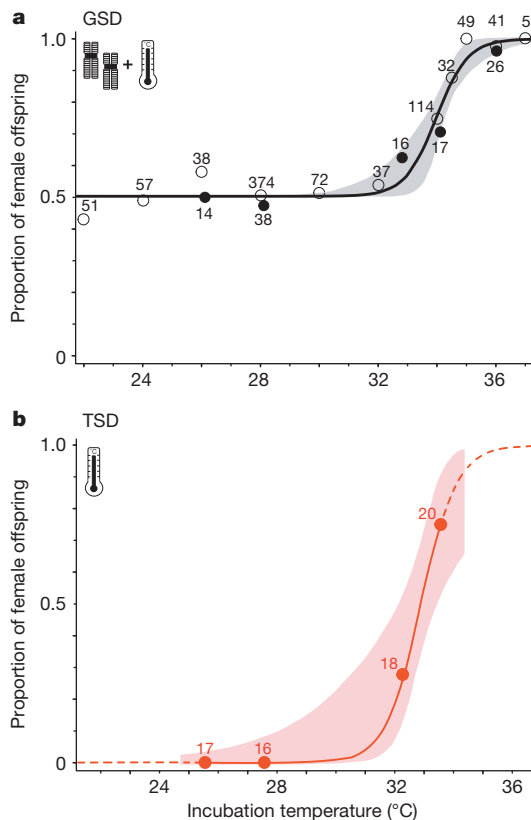


**Figure 1 | Geographical distribution of sex reversal in wild populations of *P. vitticeps*.** Location of sex-reversed ZZ females (ZZf) across years is indicated by red circles ( $N = 11$ ), normal ZW females (ZWf) by black circles ( $N = 72$ ) and normal ZZ males (ZZm) by grey circles ( $N = 48$ ). Pie charts indicate the relative proportions of ZZf and ZWf in years where sample size exceeded 15 phenotypically female individuals. The temporal trend is suggestive, but not significant ( $\chi^2 = 1.65$ , d.f. = 2,  $P = 0.44$ ).

<sup>1</sup>Institute for Applied Ecology, University of Canberra, Canberra, Australian Capital Territory 2601, Australia. <sup>2</sup>School of Life Science, La Trobe University, Melbourne, Victoria 3086, Australia.

†Present addresses: Research Center for Aquatic Genomics, National Research Institute of Fisheries Science, Fisheries Research Agency, 2-12-4 Fukuura, Kanazawa, Yokohama, Kanagawa 236-8648, Japan (K.M.); John Curtin School of Medical Research, Australian National University, Canberra, Australian Capital Territory 2601, Australia (B.A.).



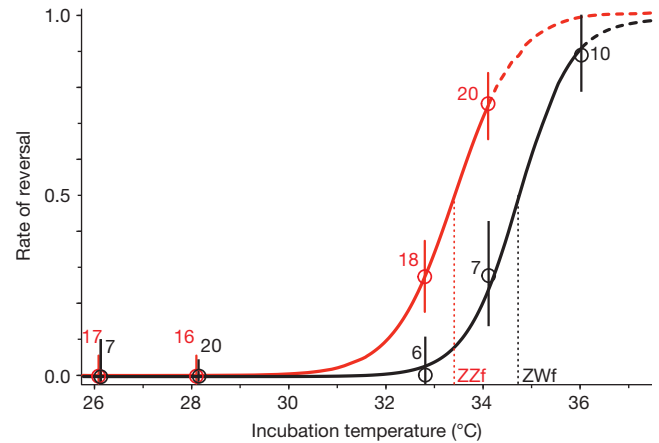


**Figure 2 | Offspring sex ratio as a function of egg incubation temperature in *P. vitticeps*.** **a**, GSD system of sex determination with a high-temperature override. Data are from ref. 12 (open circles) and this study (filled circles). Proportion of phenotypically female offspring from control ZZm × ZWf crosses is a function of constant incubation temperature  $T$ , given by  $\Pr\{f\} = 0.5033 + \frac{(1 - 0.5033)e^{1.8723T - 63.5904}}{1 + e^{1.875T - 63.5904}}$ . **b**, Functional TSD by sex reversal. Proportion of phenotypically female offspring from ZZm × ZZf crosses is given by  $\Pr\{f\} = \frac{e^{-52.9999 + 1.5866T}}{1 + e^{-52.9999 + 1.5866T}}$ . Dashed lines, extrapolation of the fitted curve beyond the data. Shaded regions, 95% confidence limits. The number of individuals in each treatment is shown.

ZZf = 47.3 eggs,  $N = 6$  females) than ZW females of an equivalent age (mean ZWf = 24.5 eggs,  $N = 11$  females) ( $P < 0.05$ ).

Four wild-caught and three captive-bred<sup>12</sup> sex-reversed females (ZZf) were mated with normal males (ZZm) under controlled laboratory conditions, to yield 389 eggs from 21 clutches. As expected from ZZm × ZZf matings, eggs incubated at the low temperature of 28 °C produced ZZ hatchlings, which were all male ( $N = 35$  hatchlings, two complete clutches, 87.5% hatching success). Parentage analysis using 2,229 single nucleotide polymorphism (SNP) markers in a subset of individuals confirmed that eggs laid by sex-reversed females were the product of sexual reproduction and not facultative parthenogenesis (Supplementary Table 1), a phenomenon that is uncommon, but known in some squamates<sup>18–20</sup>.

Following our experiments at the low incubation temperature of 28 °C, we then compared the offspring sex ratio of sex-reversed females (74 ZZf eggs) and control females (130 ZWf eggs) across a range of temperatures. Hatching success was high both for sex-reversed and for control matings (95.9% and 85.4%, respectively) (Supplementary Table 2). Offspring sex ratios from the control matings confirmed the existence of a ZW chromosomal temperature override as previously described<sup>12</sup> (Fig. 2a) (probit regression, Wald  $\chi^2 = 14.13$ , d.f. = 1,  $P < 0.0002$ ). Specifically, chromosomal influence over sex determination was dominant from 22 °C to 32 °C, where the proportion of males and females was equal (estimate of unconstrained



**Figure 3 | Rate of sex reversal as a function of egg incubation temperature in *P. vitticeps*.** Offspring of sex-reversed mothers (ZZf shown in red) are reversed more frequently and at a lower temperature than the offspring of control mothers (ZWf shown in black), implying that temperature sensitivity is variable in the population and heritable. Vertical bars, standard error of the observed proportion. Dashed lines, extrapolation of the fitted curve beyond the data (see Methods for equations). Dotted lines, the pivotal temperature at which half of ZZ offspring are reversed. Sample size (numbers) is the total number of ZZ individuals.

lower asymptote = 0.503; Fig. 2a). Temperature began to interact with and override chromosomal sex determination (causing sex reversal) above 32 °C, reaching almost complete female bias at high temperatures (proportion female at 36 °C = 0.96) (Fig. 2a; Supplementary Table 2). In contrast to the control matings, the offspring sex ratios produced by sex-reversed females followed the pattern of a TSD species with no chromosomal influence over sex (logistic regression, Wald  $\chi^2 = 8.09$ , d.f. = 1,  $P < 0.005$ ) (Fig. 2b and Supplementary Table 2). In the absence of a W chromosome, the female phenotype is possible only via sex reversal<sup>12</sup> (Fig. 3 and Supplementary Table 3). Thus, sex-reversed mothers produced only male hatchlings at low and intermediate incubation temperatures (26 °C and 28 °C), but at 33 °C produced female as well as male offspring (proportion female = 0.28) and at 34 °C produced offspring that were predominantly female (proportion female = 0.75) (Fig. 2b).

Logistic regressions of rate of sex reversal against temperature were significant for the offspring of both normal ZZm × ZWf crosses (Wald  $\chi^2 = 8.14$ , d.f. = 1,  $P < 0.005$ ) and sex-reversed ZZm × ZZf crosses (Wald  $\chi^2 = 8.09$ , d.f. = 1,  $P < 0.005$ ) (Fig. 3 and Supplementary Table 3). However, offspring of sex-reversed mothers have themselves a greater temperature sensitivity, compared with offspring of normal ZW mothers ( $\chi^2 = 55.39$ ,  $P < 0.0001$ ) (Fig. 3). The pivotal temperature for offspring of ZW mothers was 34.7 °C whereas for the offspring with sex-reversed ZZf mothers it was lower, at 33.5 °C (Fig. 3).

Our discovery of naturally occurring sex reversal has afforded us the unique opportunity to conduct ZZm × ZZf matings that can yield only ZZ offspring, whose sex is determined entirely by incubation temperature. The W chromosome was thus eliminated from our sex-reversed lineage in one generation, and sex of the offspring was determined by a mode indistinguishable from TSD. Thus, we have used high-temperature incubation to experimentally induce, within a lineage, a transition from a predominantly genotypic system with heteromorphic sex chromosomes<sup>15</sup> to a temperature-dependent sex determination system. The homomorphic sex chromosomes in ZZ male and female offspring have effectively become autosomal. Although loss of the W chromosome can be induced in a single generation under laboratory conditions, in wild populations the W chromosome is expected to persist for multiple generations and the transition in sex determination modes be more gradual.

This experimental transition from genotypic sex determination (GSD) to TSD demonstrates a novel transitional pathway, in which TSD can evolve rapidly in response to extreme environmental conditions (high temperatures), without requiring that there be sex-specific selective advantages. This observation challenges conventional theory on the drivers of transitions from GSD to TSD. Our stochastic model of the transition from GSD to TSD provides an alternative mechanism to the prevailing view that TSD evolves in response to a fitness advantage to the offspring to be female at some temperatures and male at others (invoking the Charnov–Bull model)<sup>21–23</sup>. Until now it has not been possible to disambiguate sex-specific fitness differences that could cause a GSD to TSD transition from fitness differences that arose after the transition. The Charnov–Bull model still is key in explaining the maintenance of TSD in populations; however, we show here that there are multiple pathways for populations to achieve this transition (stochastic and evolutionary). In our case, optimization of male and female fitness at the different temperatures they experience probably follows and reinforces the GSD to TSD transition, rather than being the proximal cause of the evolutionary transition.

Our demonstration that TSD can evolve rapidly in response to extreme environmental conditions (high temperatures) predicts that in the wild TSD may become fixed in small demes in which sex-reversed ZZ females mate, and the W is lost stochastically. Drift effects may be especially important if the sensitivity of ZZ males to sex reversal (the pivotal temperature for sex determination) has limited opportunity to evolve, owing to a small number of generations or low effective heritability<sup>24</sup>. Alternatively, the transition to TSD may be driven by positive selection<sup>21</sup>. Our extraordinary finding that sex-reversed females laid nearly twice as many eggs per year than normal ZW females of an equivalent age suggests an immediate fitness advantage to sex reversal<sup>25</sup> which could drive transitions. Indeed, the greater fecundity of sex-reversed females may well combine with their heritable increased propensity to reverse (Fig. 3), exacerbating the overproduction of females and accelerating the loss of the W chromosome.

In the absence of drift and positive selection, the W chromosome may still be eliminated from populations by Fisher's frequency-dependent selection<sup>3,26</sup>. An increase in the frequency of ZZ females by high-temperature reversal will be favoured by selection for mothers who produce only sons (the rarer sex at higher temperatures). Modelling this response (see Methods) shows a precipitous decline in the frequency of the ZW genotype with increasing incubation temperature. The W chromosome frequency reduces at temperatures above 32.0 °C with complete loss, enforced by frequency-dependent selection alone, at temperatures above 33.4 °C (Extended Data Fig. 4). Our wild population resides on the precipice between GSD with sex reversal and TSD arising from the loss of the W chromosome. Thus, under climatic conditions where extreme high temperatures are experienced at an increasing rate, the relatively rapid loss of the W chromosome and the adoption of TSD become increasingly likely via the combined effects of drift, positive selection and/or Fisherian sex-ratio selection. Consistent with these predictions we observed increased rates of sex reversal among wild phenotypic females over the course of this study (Fig. 1).

Our observation that offspring from sex-reversed mothers are more frequently reversed and at a lower temperature than the offspring of normal ZW mothers (Fig. 3) implies that the eggs of ZZ sex-reversed females are more sensitive to temperature. This strongly suggests that heritable variation (genetic or epigenetic) exists at the locus that controls sex determination in *P. vitticeps*, providing a mechanism for sex determination thresholds to evolve. Additional evidence for heritable selectable variation in thermosensitivity was observed in a single sex-reversed female that produced 100% male offspring at all temperatures (28–36 °C; *N* = 32 eggs; Supplementary Tables 2 and 3; data excluded from analyses).

One of the most important questions for the near future is whether organisms will be sufficiently resilient to withstand a rapidly changing

climate, or so vulnerable that they will succumb to extinction<sup>27,28</sup>. We provide here an example of how climatic extremes can rapidly and fundamentally alter the biology (switches in sex determination mode) and the genome (loss of the W chromosome) of climate-sensitive reptiles. Thus adverse evolutionary responses, specifically a switch to TSD if global temperatures rise, can occur through a combination of temperature sensitivity and stochastic processes. This is important because temperature-induced extreme sex-ratio bias is thought to be an extinction driver exclusively in species with TSD<sup>29</sup>. However, we show here that these risks may extend more broadly. Exposure to high temperatures can perturb apparently stable GSD systems, induce a rapid transition to TSD and then proceed inexorably towards a highly feminized population and thus a greater risk of extinction.

The key to determining how important transitions in sex-determining mode are to species' survival will be to discover how thermal sensitivities can adapt after the transition to TSD. If adaptation is rapid, TSD could produce stable unbiased sex ratios, and may even be favoured by climate change. For example, a temperature-dependent strategy might afford greater control of sex ratio manipulation in an unpredictable climate. In this way, reptiles may have greater capacity to cope and compensate for climate change than previously appreciated<sup>27</sup>. A high degree of flexibility in sex-determination mode could be a powerful and, until now, unappreciated weapon in the arsenal of evolutionary responses to an unpredictable climate.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 25 February; accepted 18 May 2015.**

1. Sarre, S. D., Georges, A. & Quinn, A. The ends of a continuum: genetic and temperature-dependent sex determination in reptiles. *Bioessays* **26**, 639–645 (2004).
2. Sarre, S. D., Ezaz, T. & Georges, A. Transitions between sex-determining systems in reptiles and amphibians. *Annu. Rev. Genom. Hum. Genet.* **12**, 391–406 (2011).
3. Bull, J. J. Evolution of environmental sex determination from genotypic sex determination. *Heredity* **47**, 173–184 (1981).
4. Quinn, A. E., Sarre, S. D., Ezaz, T., Graves, J. A. M. & Georges, A. Evolutionary transitions between mechanisms of sex determination in vertebrates. *Biol. Lett.* **7**, 443–448 (2011).
5. Schwanz, L. E., Ezaz, T., Gruber, B. & Georges, A. Novel evolutionary pathways of sex-determining mechanisms. *J. Evol. Biol.* **26**, 2544–2557 (2013).
6. Chan, S. T. H. & Yeung, W. S. B. Sex control and sex reversal in fish under natural conditions. *Fish Physiol.* **9**, 171–222 (1983).
7. Alho, J. S., Matsuba, C. & Merila, J. Sex reversal and primary sex ratios in the common frog (*Rana temporaria*). *Mol. Ecol.* **19**, 1763–1773 (2010).
8. Miura, I. Sex chromosome differentiation in the Japanese brown frog, *Rana japonica*. I. Sex-related heteromorphism of the distribution pattern of constitutive heterochromatin in chromosome no. 4 of the Wakuya population. *Zool. Sci.* **11**, 797–806 (1994).
9. Bachtrog, D. *et al.* Sex determination: why so many ways of doing it? *PLoS Biol.* **12**, e1001899 (2014).
10. Shine, R., Elphick, M. J. & Donnellan, S. Co-occurrence of multiple, supposedly incompatible modes of sex determination in a lizard population. *Ecol. Lett.* **5**, 486–489 (2002).
11. Radder, R. S., Quinn, A. E., Georges, A., Sarre, S. D. & Shine, R. Genetic evidence for co-occurrence of chromosomal and thermal sex-determining systems in a lizard. *Biol. Lett.* **4**, 176–178 (2008).
12. Quinn, A. E. *et al.* Temperature sex reversal implies sex gene dosage in a reptile. *Science* **316**, 411 (2007).
13. Mork, L., Czerwinski, M. & Capel, B. Predetermination of sexual fate in a turtle with temperature-dependent sex determination. *Dev. Biol.* **386**, 264–271 (2014).
14. Wang, C. *et al.* Identification of sex chromosomes by means of comparative genomic hybridization in a lizard, *Eremias multiocellata*. *Zool. Sci.* **32**, 151–156 (2015).
15. Ezaz, T. *et al.* The dragon lizard *Pogona vitticeps* has ZZ/ZW micro-sex chromosomes. *Chromosome Res.* **13**, 763–776 (2005).
16. Quinn, A. E., Ezaz, T., Sarre, S. D., Graves, J. A. M. & Georges, A. Extension, single-locus conversion and physical mapping of sex chromosome sequences identify the Z microchromosome and pseudo-autosomal region in a dragon lizard, *Pogona vitticeps*. *Heredity* **104**, 410–417 (2010).
17. Ezaz, T. *et al.* Sequence and gene content of a large fragment of a lizard sex chromosome and evaluation of candidate sex differentiating gene R-spondin 1. *BMC Genom.* **14**, 899 (2013).
18. Booth, W. *et al.* Facultative parthenogenesis discovered in wild vertebrates. *Biol. Lett.* **8**, 983–985 (2012).
19. Booth, W., Johnson, D. H., Moore, S., Schal, C. & Vargo, E. L. Evidence for viable, non-clonal but fatherless boa constrictors. *Biol. Lett.* **7**, 253–256 (2011).

20. Watts, P. C. *et al.* Parthenogenesis in Komodo dragons. *Nature* **444**, 1021–1022 (2006).
21. Charnov, E. L. & Bull, J. When is sex environmentally determined? *Nature* **266**, 828–830 (1977).
22. Warner, D. A. & Shine, R. The adaptive significance of temperature-dependent sex determination in a reptile. *Nature* **451**, 566–568 (2008).
23. Warner, D. A. & Shine, R. The adaptive significance of temperature-dependent sex determination: experimental tests with a short lived lizard. *Evolution* **59**, 2209–2221 (2005).
24. McGaugh, S. E. & Janzen, F. J. Effective heritability of targets of sex-ratio selection under environmental sex determination. *J. Evol. Biol.* **24**, 784–794 (2011).
25. Saunders, P. A. *et al.* XY females do better than the XX in the african pygmy mouse, *Mus minutoides*. *Evolution* **68**, 2119–2127 (2014).
26. Bull, J. J. *Evolution of Sex Determining Mechanisms* (Benjamin-Cummings, 1983).
27. Sinervo, B. *et al.* Erosion of lizard diversity by climate change and altered thermal niches. *Science* **328**, 894–899 (2010).
28. Hoffmann, A. A. & Sgro, C. M. Climate change and evolutionary adaptation. *Nature* **470**, 479–485 (2011).
29. Boyle, M., Hone, J., Schwanz, L. E. & Georges, A. Under what conditions do climate-driven sex ratios enhance versus diminish population persistence? *Ecol. Evol.* **4**, 4522–4533 (2014).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank the following persons and institutions: W. Ruscoe and J. Richardson (animal husbandry); A. Quinn, M. Young and J. Richardson (field

collections); A. Kilian and Diversity Arrays Technology (SNP genotyping); A. Livernois (sequencing assistance); B. Gruber and N. Garlapati (geographic information system); A. T. Adamack (modelling advice); A. Dobos and P. E. Geertz (graphic design); J. Deakin, R. Thompson and the Kioloa Science Writers Workshop (revisions of the manuscript). We particularly thank reviewer J. J. Bull for suggesting the modelling approach in Extended Data Fig. 4. Funding was from Australian Research Council Discovery Grant DP110104377 to A.G. and T.E. This research was conducted under appropriate approvals from the Victorian, New South Wales and Queensland authorities, and with approvals from the Animal Ethics Committee of the University of Canberra.

**Author Contributions** C.E.H. and A.G. designed the study. C.E.H. conducted breeding experiments, egg incubations, parentage SNP analysis and prepared figures. D.O'M. collected the animals from the field. C.E.H. and X.Z. conducted the molecular sex testing. B.A. and K.M. undertook the cytogenetic analysis and prepared extended data figures, under the supervision of T.E. A.G. and C.E.H. undertook the statistical analyses and A.G. conducted the modelling of ZW genotype frequency with temperature. All authors contributed to writing the manuscript.

**Author Information** The W chromosome sequence has been deposited in GenBank under accession number KM508988. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.E.H. ([clare@holleley.net](mailto:clare@holleley.net)) or A.G. ([georges@aerg.canberra.edu.au](mailto:georges@aerg.canberra.edu.au)).



## METHODS

**Field collection.** We collected samples (tail snips or blood) from 131 wild adult *P. vitticeps* individuals, during sampling trips conducted in October 2003, 2004, 2008, September 2009 and March 2010, 2011. Phenotypic sex was determined by the presence or absence of hemipenes visible as two latero-ventral lumps on the tail immediately behind the vent and subsequent eversion of hemipenes if present, and presence or absence of eggs by palpation, with confirmatory but not definitive presence or absence of enlarged femoral pores and larger head size characteristic of males.

**Breeding experiments.** Captive females were allowed to lay eggs naturally in a sand substrate. Eggs were recovered from cages within 14 h and transferred to plastic boxes filled with vermiculite with a water potential of  $-200$  kPa (120.0% water to vermiculite, by mass)<sup>23</sup>. Allocation of clutches to experiments was not randomized and the investigators were not blinded to allocation during experiments and outcome assessment. In the first experiment, two whole clutches (21 and 19 eggs) from a sex-reversed female (ZZf) were selected as they became available and incubated at a constant  $28^{\circ}\text{C}$  ( $28.1^{\circ}\text{C}$ , s.d.  $\pm 0.6$ ) to evaluate reproductive viability. In the second experiment, clutches laid both by sex-reversed (ZZf) and by control (ZWf) females were chosen randomly then systematically allocated across four constant temperature incubation treatments:  $26^{\circ}\text{C}$ ,  $28^{\circ}\text{C}$ ,  $33^{\circ}\text{C}$ ,  $34^{\circ}\text{C}$  (Supplementary Table 2). An additional  $36^{\circ}\text{C}$  treatment was conducted only for the eggs of control females because of lower numbers of available eggs from ZZf. Sample sizes were determined by availability of clutches from the captive breeding programme. No statistical methods were used to predetermine sample size. We have not interpreted non-significant results, where considerations of power are more crucial. Eggs were removed from the experiment if they were physically damaged (crushed or pierced) or inviable (no white patch or vascularization) before entering the thermosensitive period when sex is determined<sup>30</sup> (Supplementary Table 4). Whole clutches with egg mortality rates of at least 40% were excluded from analysis. The phenotypic sex for all captive-bred animals was established by hemipene eversion upon hatching<sup>31</sup>, hemipenal transillumination<sup>32</sup> at 1–4 months and by gross external morphology at 4–9 months of age. Female reproductive fitness was estimated as the total number of eggs produced per season, a measure that summarizes the effects of clutching frequency and clutch size. We compared the reproductive fitness of 6 sex-reversed (ZZf) and 12 control (ZWf) females using unpaired *t*-tests.

**Molecular detection of sex reversal.** All individuals collected from the wild and bred in captivity were genotypically sexed using two PCR primers: H2, GCCCATATCTCACTAGTCCCTCC; F, CAGTTCCTTCTACCTGGGAGT GC, which flank two W-chromosome-specific deletions (150 base pairs and 14 base pairs) that we identified in published *P. vitticeps* anonymous sex chromosome sequence<sup>16</sup> (GenBank accession numbers EU938138.1 and KM508988). PCR conditions for the novel test were  $1\times$  MyTaq HS Red mix (Bioline),  $4\mu\text{M}$  each primer and 50 ng of genomic DNA. Cycling conditions were  $95^{\circ}\text{C}$  for 5 min; ( $95^{\circ}\text{C}$  for 20 s,  $70$ – $65^{\circ}\text{C}$  for 20 s,  $72^{\circ}\text{C}$  for 1 min)  $\times$  ten cycles with annealing temperature decreased  $0.5^{\circ}\text{C}$  per cycle; ( $95^{\circ}\text{C}$  for 20 s,  $65^{\circ}\text{C}$  for 20 s,  $72^{\circ}\text{C}$  for 1 min)  $\times$  30 cycles;  $72^{\circ}\text{C}$  for 10 min. PCR products were visualized on a 1.5% agarose gel using SYBR Safe (Life Technologies). Two bands amplified in ZW individuals, whereas a single control band amplified in ZZ individuals. Individuals showing genotype–phenotype discordance were classed as sex-reversed. The rate of sex reversal was calculated as the proportion of ZZ individuals with a female phenotype. All molecular sex tests were conducted with the investigator blinded to the identity and phenotypic sex of the samples.

**Cytogenetics.** The accuracy of the PCR sex test was validated using C-banding (Extended Data Fig. 1), comparative genomic hybridization (Extended Data Fig. 2) and by physically mapping a *P. vitticeps* W-chromosome-linked microsatellite motif (Extended Data Fig. 3). Two wild-type ZW females (001003386049, 001003342236), two wild-type ZZ males (001003338787, 001003387339) and a putative sex reversal female (001003344224) of *P. vitticeps* were used for cytogenetic analyses. Metaphase chromosome spreads were prepared from fibroblast cultures of tail tissue following ref. 33. Metaphases for all individuals were stained with 4',6-diamidino-2-phenylindole (DAPI), the chromosome number identified and compared with the normal *P. vitticeps* karyotype<sup>34</sup>, to eliminate the possibility of chromosome abnormality. C-banded chromosomes were obtained by the CBG method (C-bands by barium hydroxide using Giemsa)<sup>15,35</sup>. Comparative genomic hybridization was conducted as previously described<sup>15,36</sup> using fluorescently labelled male and female genomic DNA. Physical mapping of the W-chromosome-linked microsatellite (AAGG)<sub>8</sub> was conducted using fluorescence *in situ* hybridization, following ref. 37. For all cytogenetic analyses, the presence or absence of W-chromosome-specific signal was scored in eight to ten metaphases per individual.

**Statistical analysis.** Curves of best-fit offspring responses to incubation temperature were estimated by applying logistic regressions to the raw data (male = 0;

female = 1) using PROC LOGISTIC in SAS Software version 9.1 or, in the case of the proportion of females varying from 0.5 to 1, by fitting a logistic regression with unconstrained lower asymptote using PROC PROBIT. In both cases, the SAS ODS output was used to generate 95% confidence limits around the estimated regression lines. Logistic regressions were compared using PROC LOGISTIC with the addition of a CLASS variable for maternal genotype (ZZ  $\times$  ZZ versus ZZ  $\times$  ZW). These analyses have an appropriate error structure for data in the form of counts of males and females, and counts of sex-reversed versus concordant individuals.

**Parentage analysis.** Parentage analysis was conducted to confirm that the eggs produced by sex-reversed females (ZZf) were the product of sexual reproduction and not asexual reproduction by parthenogenesis. We genotyped 18 individuals at 2,229 SNPs using a proprietary reduced-representation sequencing approach called DArTseq (Diversity Arrays Technology), alternatively referred to as double-digest restriction-site-associated DNA markers (RAD-seq)<sup>38–41</sup>. Four methods of complexity reduction were tested in *P. vitticeps* (data not presented) and the PstI–SphI method was selected. Approximately 2,000,000 sequences per barcode per sample were identified and used in marker calling. We subsequently applied further stringent quality control measures, requiring 100% reproducibility over two independent runs, a minimum of  $5\times$  read depth (mean read depth =  $32\times$ ) and complete data for all individuals. Specifically, we sequenced three parents (two sex-reversed females mated to the same male), nine offspring from the first pairing, two offspring from the second pairing, three unrelated ZW individuals and one unrelated ZZ individual. We used these data to estimate mother–offspring sequence identity (clonal parthenogenesis hypothesis), the percentage of heterozygous loci in offspring (non-clonal parthenogenesis hypothesis) and the percentage of parent–offspring allelic mismatches (sexual reproduction hypothesis). For comparison, these statistics were also calculated comparing four unrelated individuals to the parental genotypes. Parentage hypotheses were tested using *t*-tests (Supplementary Table 1).

**Modelling the decline of the ZW genotype.** We modelled the decline of the ZW genotype resulting from frequency-dependent selection because of overproduction of females through sex reversal with increasing temperature<sup>3</sup>. Fitness within a sex is the same for all genotypes. Let the starting frequency of ZW among zygotes be  $y$ , the starting frequency of ZZ be  $z = 1 - y$ , and let a fraction  $P_1[T]$  of ZZ become reversed to a female phenotype if they have a ZW mother,  $P_2[T]$  if they have a ZZ mother. The equations for  $P_1[T]$  and  $P_2[T]$  are the functions of temperature  $T$  given in Fig. 3. In any given generation  $n$ , we have the proportion of female phenotypes ( $f_n$ ) equal to the sum of the numbers of normal ZW females ( $y_n$ ), of sex-reversed ZZ females with ZW mothers ( $r_n$ ) and of sex-reversed ZZ females with ZZ mothers ( $r'_n$ ),

$$f_n = y_n + r_n + r'_n$$

The frequency of ZW zygotes and the frequency of ZZ zygotes with ZW mothers are equal and both given by

$$y_{n+1} = z_{n+1} = \frac{y_n}{2f_n}$$

The frequency of sex-reversed ZZ zygotes with ZW mothers thus given by

$$r_{n+1} = P_1[T]z_{n+1}$$

and the frequency of sex-reversed ZZ zygotes with ZZ mothers is given by

$$r'_{n+1} = P_2[T]\frac{(f_n - y_n)}{f_n}$$

Determined from our experimental data, the probability of sex reversal for offspring with ZW mother (Fig. 3, shown in black) is given by

$$P_1[T] = \frac{e^{-63.1402 + 1.8184T}}{1 + e^{-63.1402 + 1.8184T}}$$

and the probability of sex reversal for offspring with ZZ mothers (Fig. 3, shown in red) is given by

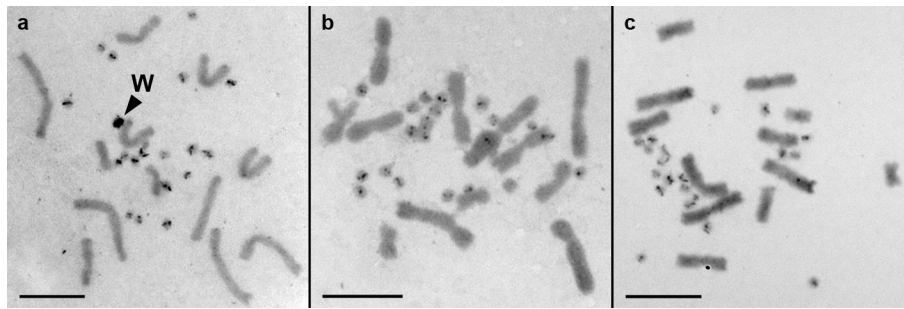
$$P_2[T] = \frac{e^{-52.9999 + 1.5866T}}{1 + e^{-52.9999 + 1.5866T}}$$

We iterate for an equilibrium solution for  $y$  for various values of temperature  $T$ . Overlapping generations will delay the rate of convergence to equilibrium, but will not affect the equilibrium value for a particular temperature.

30. Mrosovsky, N. & Pieau, C. Transitional range of temperature, pivotal temperatures and thermosensitive stages for sex determination in reptiles. *Amphib.-Reptil.* **12**, 169–179 (1991).

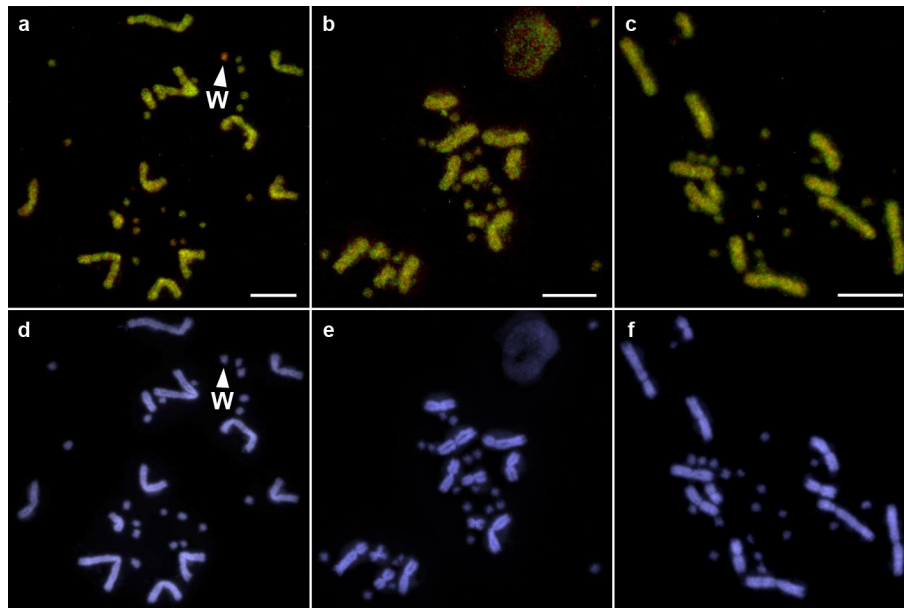
31. Harlow, P. S. A harmless technique for sexing hatchling lizards. *Herpetol. Rev.* **27**, 71–72 (1996).

32. Brown, D. Hemipenial transillumination as a sexing technique in varanids. *Biawak* **3**, 26–29 (2009).
33. Ezaz, T. *et al.* A simple non-invasive protocol to establish primary cell lines from tail and toe explants for cytogenetic studies in Australian dragon lizards (Squamata: Agamidae). *Cytotechnology* **58**, 135–139 (2008).
34. Young, M. J., O'Meally, D., Sarre, S. D., Georges, A. & Ezaz, T. Molecular cytogenetic map of the central bearded dragon, *Pogona vitticeps* (Squamata: Agamidae). *Chromosome Res.* **21**, 361–374 (2013).
35. Sumner, A. T. A simple technique for demonstrating centromeric heterochromatin. *Exp. Cell Res.* **75**, 304–306 (1972).
36. Matsubara, K. *et al.* Highly differentiated ZW sex microchromosomes in the Australian *Varanus* species evolved through rapid amplification of repetitive sequences. *PLoS ONE* **9**, e95226 (2014).
37. Cioffi, M. B., Martins, C., Vicari, M. R., Rebordinos, L. & Bertollo, L. A. Differentiation of the XY sex chromosomes in the fish *Hoplias malabaricus* (Characiformes, Erythrinidae): unusual accumulation of repetitive sequences on the X chromosome. *Sex Dev.* **4**, 176–185 (2010).
38. Courtois, B. *et al.* Genome-wide association mapping of root traits in a japonica rice panel. *PLoS ONE* **8**, e78037 (2013).
39. Cruz, V. M., Kilian, A. & Dierig, D. A. Development of DArT marker platforms and genetic diversity assessment of the U.S. collection of the new oilseed crop lesquerella and related species. *PLoS ONE* **8**, e64062 (2013).
40. Kilian, A. *et al.* Diversity arrays technology: a generic genome profiling technology on open platforms. *Methods Mol. Biol.* **888**, 67–89 (2012).
41. Raman, H. *et al.* Genome-wide delineation of natural variation for pod shatter resistance in *Brassica napus*. *PLoS ONE* **9**, e101673 (2014).



**Extended Data Figure 1** | C-banded *P. vitticeps* chromosomes. **a**, Mitotic metaphase chromosomes of a ZW control female individual. Arrowhead indicates the presence of a W chromosome identified by dense black staining of a single microchromosome. **b**, Mitotic metaphase chromosomes of a female putative ZZ sex-reversed individual. No evidence of a W chromosome was detected. **c**, Mitotic metaphase chromosomes of a control ZZ male individual. No evidence of a W chromosome was detected. Scale bar, 10  $\mu\text{m}$ .





### Extended Data Figure 2 | Comparative genomic hybridization in

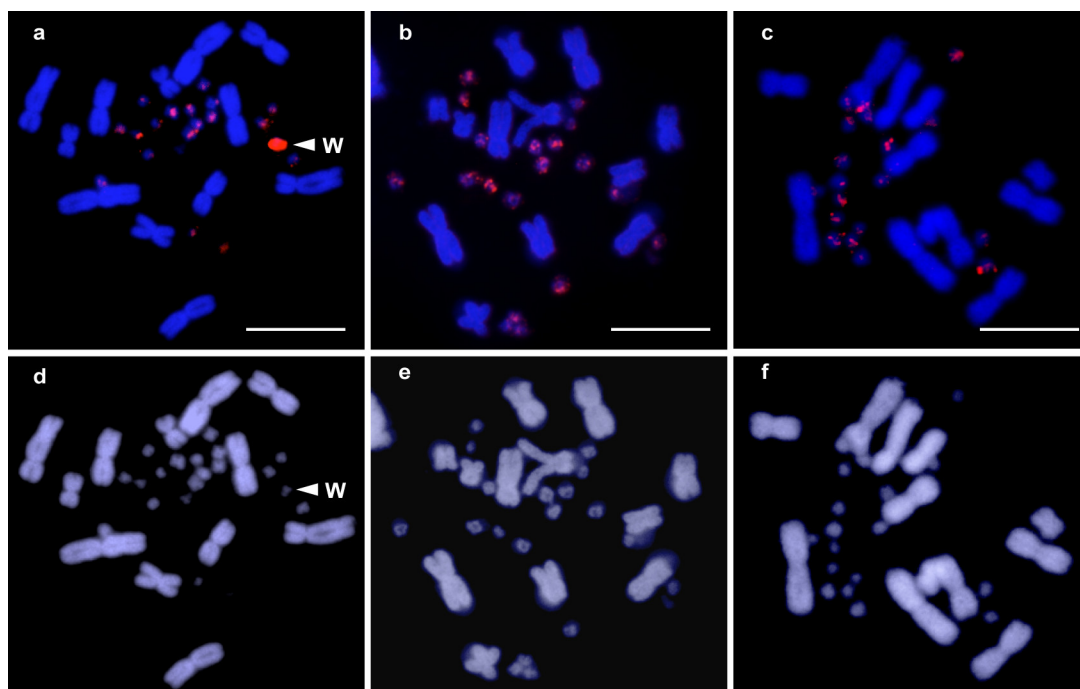
*P. vitticeps*. Genomic DNA was labelled by nick translation incorporating SpectrumGreen-dUTP for males and SpectrumOrange-dUTP for females.

**a**, Mitotic metaphase chromosomes of a ZW control female individual.

Arrowhead indicates the presence of a single W microchromosome identified by the enriched orange fluorescence of female specific genomic DNA labelled

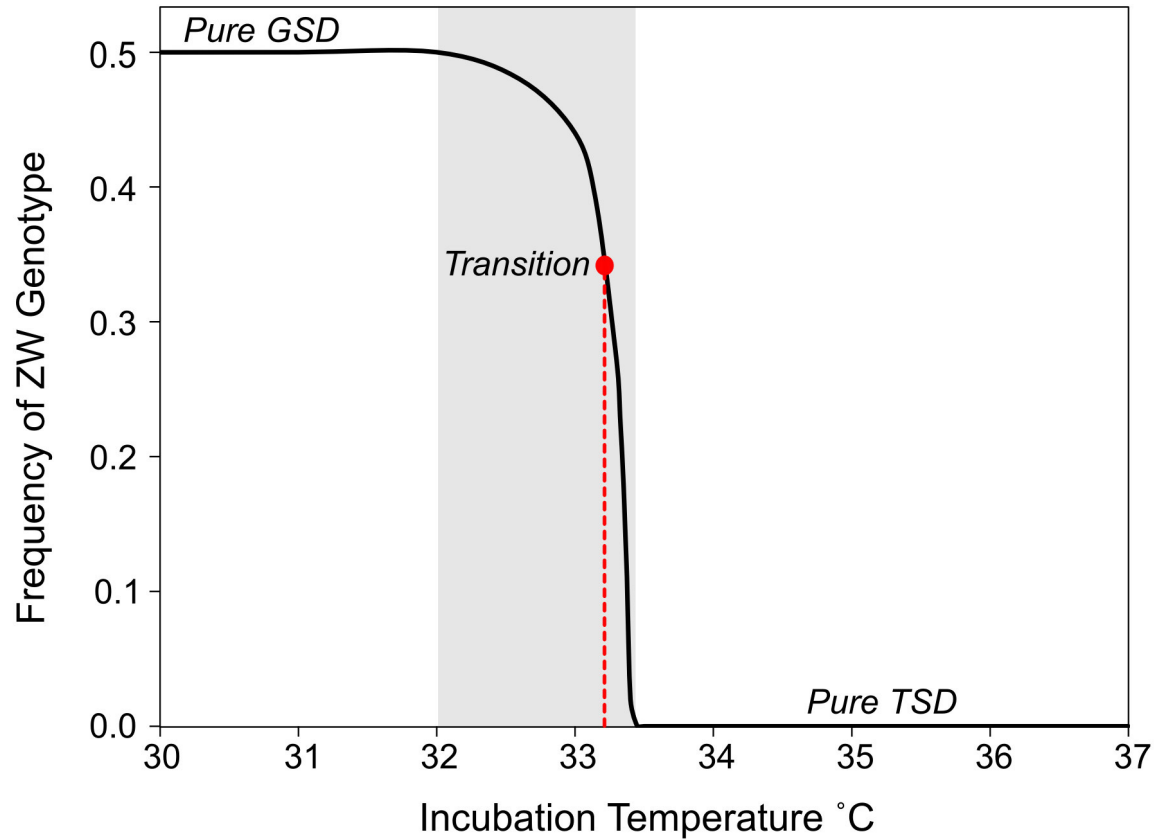
with SpectrumOrange-dUTP. **b**, Mitotic metaphase chromosomes of a female putative ZZ sex-reversed individual. No evidence of a W chromosome was detected. **c**, Mitotic metaphase chromosomes of a control ZZ male individual. No evidence of a W chromosome was detected.

**d–f**, DAPI staining of the same metaphases, control ZW female, sex-reversed ZZ female and control ZZ male, respectively. Scale bar, 10  $\mu$ m.



**Extended Data Figure 3 | Physical mapping of a W-chromosome-linked microsatellite motif in *P. vitticeps*.** **a**, Mitotic metaphase chromosomes of a ZW control female individual. Arrowhead indicates the presence of a W chromosome identified by a strong hybridization of (AAGG)<sub>8</sub>-Cy3 fluorescence (orange) on a single microchromosome. **b**, Mitotic metaphase chromosomes of

a female putative ZZ sex-reversed individual. No evidence of a W chromosome was detected. **c**, Mitotic metaphase chromosomes of a control ZZ male individual. No evidence of a W chromosome was detected. **d–f**, DAPI staining of the same metaphases, control ZW female, sex-reversed ZZ female and control ZZ male, respectively. Scale bar, 10  $\mu$ m.



**Extended Data Figure 4 | Modelling the decline of the ZW genotype resulting from frequency-dependent selection.** Frequency of the ZW genotype declines precipitously with increasing incubation temperature. Our

wild population (shown in red, 14.3% sex reversal) resides on the precipice between GSD and TSD and requires only a small change in environmental temperature to precipitate loss of the W chromosome.



# Spatiotemporal control of a novel synaptic organizer molecule

Kelly Howell<sup>1</sup>, John G. White<sup>2</sup> & Oliver Hobert<sup>1</sup>

**Synapse formation is a process tightly controlled in space and time. How gene regulatory mechanisms specify spatial and temporal aspects of synapse formation is not well understood. In the nematode *Caenorhabditis elegans*, two subtypes of the D-type inhibitory motor neuron (MN) classes, the dorsal D (DD) and ventral D (VD) neurons, extend axons along both the dorsal and ventral nerve cords<sup>1</sup>. The embryonically generated DD motor neurons initially innervate ventral muscles in the first (L1) larval stage and receive their synaptic input from cholinergic motor neurons in the dorsal cord. They rewire by the end of the L1 moult to innervate dorsal muscles and to be innervated by newly formed ventral cholinergic motor neurons<sup>1</sup>. VD motor neurons develop after the L1 moult; they take over the innervation of ventral muscles and receive their synaptic input from dorsal cholinergic motor neurons. We show here that the spatiotemporal control of synaptic wiring of the D-type neurons is controlled by an intersectional transcriptional strategy in which the UNC-30 Pitx-type homeodomain transcription factor acts together, in embryonic and early larval stages, with the temporally controlled LIN-14 transcription factor to prevent premature synapse rewiring of the DD motor neurons and, together with the UNC-55 nuclear hormone receptor, to prevent aberrant VD synaptic wiring in later larval and adult stages. A key effector of this intersectional transcription factor combination is a novel synaptic organizer molecule, the single immunoglobulin domain protein OIG-1. OIG-1 is perisynaptically localized along the synaptic outputs of the D-type motor neurons in a temporally controlled manner and is required for appropriate selection of both pre- and post-synaptic partners.**

At the end of the first larval stage, the synaptic outputs from the DD motor neurons (MNs) to ventral muscle and their synaptic input from cholinergic DA/DB MNs is eliminated and, instead, synapses are formed onto dorsal muscle and synaptic input is received from cholinergic VA/VB MNs (Fig. 1a)<sup>1</sup>. We sought to examine how the spatiotemporal specificity of this rewiring process is controlled and integrated with other aspects of D-type MN differentiation. To address this question, we examined the function of the *C. elegans* Pitx-type homeobox transcription factor UNC-30, which is known to control GABAergic neurotransmitter identity of the D-type MNs<sup>2,3</sup>. The analysis of serial electron micrographs shows that the synaptic patterns of the DD and VD neurons are substantially disrupted in *unc-30* null mutant animals. In adult *unc-30(0)* animals, VD MNs display ectopic synapses onto dorsal muscle and lack notable synaptic inputs from DA/DB on the dorsal side (Fig. 1b and Extended Data Fig. 1). Furthermore, DD MNs, which normally only form synapses onto ventral muscle, show aberrant innervation of dorsal muscle in L1 stage *unc-30* mutants (Fig. 1c). These synaptic defects were confirmed with green fluorescent protein (GFP)-tagged RAB-3 protein, expressed specifically in D-type MNs (Fig. 1d).

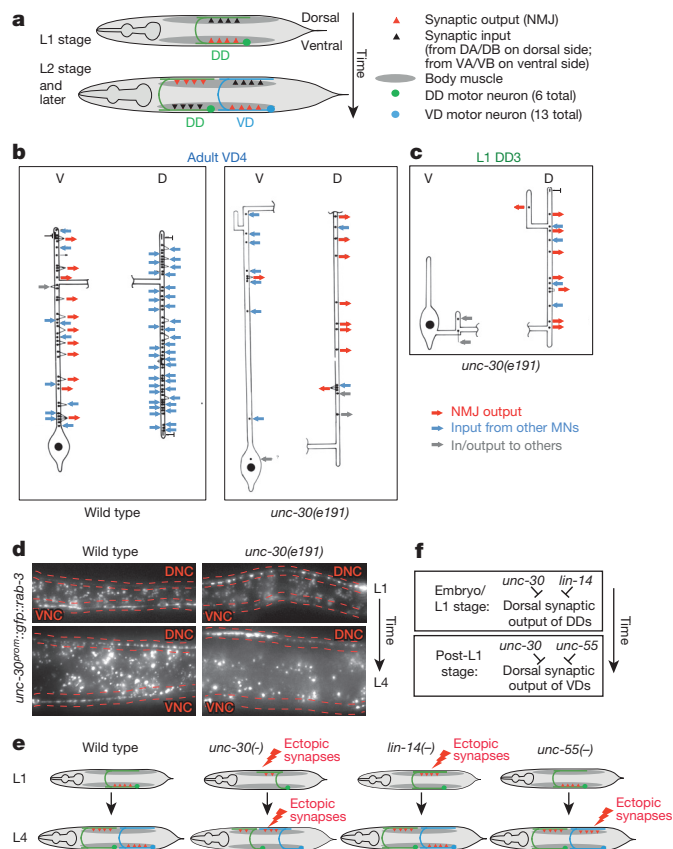
*unc-30* is expressed in both DD and VD MNs at all stages<sup>3</sup>, yet *unc-30* inhibits dorsal DD synapses only in the L1 stage and not at later stages. However, at these later stages, *unc-30* does inhibit dorsal

synapses from VD neurons, but not the DD neurons. How can the temporal and spatial specificity of the *unc-30(e191)* defects be explained? A potential answer to this question lies in the previously described mutant phenotype of two transcription factors, which recapitulate specific components of the cell-type specific, DD and VD synaptogenic defects of *unc-30* mutants. In animals lacking the *lin-14* transcription factor, whose expression is normally temporally restricted to embryonic and first larval stages in most tissues, including the D-type motoneurons<sup>4,5</sup>, DD MNs form ectopic synapses in the dorsal cord in embryonic and L1 stages (Extended Data Fig. 2a; schematized in Fig. 1e, f)<sup>4</sup>. These DD MN defects are similar to those that we observe in *unc-30* mutants. The dorsal ectopic synapses in the VD neurons of *unc-30* mutant animals (not observed in *lin-14*) are in turn recapitulated in animals lacking the *unc-55* orphan nuclear receptor, in which VD MNs form aberrant synapses in the dorsal cord, as previously shown (Extended Data Fig. 2b; schematized in Fig. 1e, f)<sup>6,7</sup>, while DD wiring at the L1 stage is normal. Taken together, the *unc-30* phenotype in the DD and VD neurons can be viewed as a 'composite' of the two individual phenotypes of *lin-14* (DD neurons at L1 stage) and *unc-55* (VD neurons at later stages) (Fig. 1e, f schematic). One possible way to explain these concordances of phenotypes is that *unc-30* may collaborate with *lin-14* to control the expression of a molecule that acts in a temporally restricted manner in embryonic and L1 stages to inhibit dorsal synapse formation of the DD neurons. In the VD MNs, *unc-30* may in turn collaborate with *unc-55* to control expression of a molecule that acts in the VD neurons to inhibit dorsal synapse formation of these neurons.

We sought to identify such potential effector molecule(s) through a candidate gene approach. In a survey of *C. elegans* immunoglobulin superfamily members, we had previously described a family of small proteins that are composed of a single Ig domain, the *oig* gene family<sup>8</sup>. One of the *oig* family members, *oig-1*, encodes a 137 amino acid-long protein with a signal sequence and a single IgC2-type domain, but no transmembrane domain or predicted glycosylphosphatidylinositol (GPI) anchor. Transgenic animals carrying an *oig-1* fosmid-based reporter construct showed expression both in the DD and VD MNs, but no other ventral nerve cord MNs (Fig. 2a). Notably, expression of *oig-1* in the D-type MNs is temporally controlled in a manner that correlates with the distinct periods of inhibition of dorsal muscle innervation exhibited by DD and VD neurons. *oig-1* is transiently expressed in the DD neurons during the time when no dorsal synapses are formed (embryos and L1), but is downregulated in the DD neurons upon formation of their dorsal synapses (L2 and later; Fig. 2a). In contrast, expression of *oig-1* in the VD neurons, which have processes but no synaptic outputs in the dorsal cord, is continuously maintained throughout the life of the neuron (Fig. 2a).

The transient expression in the DD and continuous expression in the VD neurons makes *oig-1* a candidate effector gene for the *unc-30*, *lin-14* and *unc-55* transcription factors. Indeed, the *oig-1<sup>fosmid</sup>::sl2::gfp* reporter fails to be expressed in both DD- and VD-type MNs in *unc-30* null mutants at all stages (Fig. 2b). In *lin-14* null mutant animals,

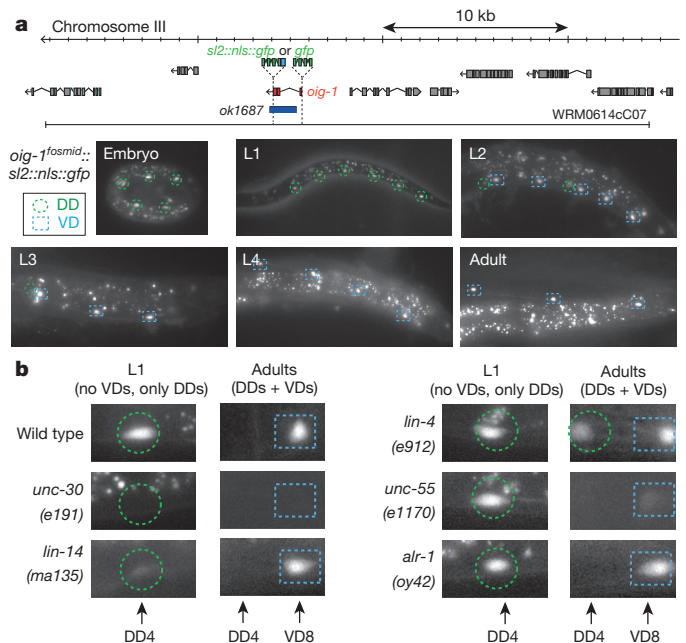
<sup>1</sup>Department of Biochemistry and Molecular Biophysics, Howard Hughes Medical Institute, Columbia University Medical Center, New York, New York 10032, USA. <sup>2</sup>MRC Laboratory of Molecular Biology, Cambridge CB2 2QH, UK.



**Figure 1 | Loss of *unc-30* disrupts the synaptic connectivity of the DD and VD MNs.** **a**, Schematic of DD rewiring<sup>1</sup>. **b**, Reconstruction of a VD4 MN from an *unc-30(e191)* adult animal compared to the same neuron in a wild-type animal. Extended Data Fig. 1 shows a more detailed presentation of the EM data. **c**, Reconstructed DD3 neuron from an *unc-30(e191)* L1 larva showing aberrant NMJs in the dorsal cord (D). Previous reconstructions of a wild-type L1 using the same techniques and personnel showed that a reconstructed DD3 made no NMJs on dorsal muscles and 9 NMJs on ventral muscles<sup>1</sup>. **d**, Presynaptic marker RAB-3 ectopically localizes mostly to ventral cord in wild-type L1 animals (16/20 animals), but ectopically in the dorsal nerve cord (DNC; outlined in red) in *unc-30* mutant animals (19/20). At the L4 stage, in which presynaptic specialization is observed in both ventral nerve cord (VNC; outlined in red) and DNC (19/20 animals), *unc-30* mutants show few specializations in the VNC (20/20 animals). Original magnification,  $\times 630$ . **e**, **f**, Summary of synapse formation defects in *unc-30*, *lin-14*, and *unc-55* mutants (**e**) and genetic interpretation (**f**).

transient expression of the *oig-1* reporter in the L1 stage is diminished in the DD MNs (Fig. 2b) and temporally prolonged expression of *lin-14*, achieved through genetic removal of a negative regulator of *lin-14*, the microRNA *lin-4* (ref. 9), results in prolonged expression of *oig-1* in the DD MNs into the adult stage (Fig. 2b). In animals lacking *unc-55*, which is normally expressed in VD, but not DD MNs<sup>10</sup>, expression of *oig-1* in the DD neurons at the L1 stage is unaffected, but *oig-1* expression in the VD neurons is absent at the adult stage (Fig. 2b). The two distinct transcription factor combinations that control *oig-1* expression in DD (*unc-30* and *lin-14*) and in VD (*unc-30* and *unc-55*) operate independently since the expression of each transcription factor is independent of the presence of the other transcription factor (Extended Data Fig. 3a, b)<sup>11</sup>.

The transcriptional nature of *oig-1* regulation is corroborated by the finding that 1 kilobase (kb) of 5' sequences of *oig-1* conveys the same spatiotemporal regulation as the *oig-1* fosmid reporter (Extended Data Fig. 4). Chromatin immunoprecipitation-sequencing data from the modEncode project shows binding of UNC-55 to this 1-kb fragment of the *oig-1* locus<sup>12</sup>, suggesting direct regulation. However, it is

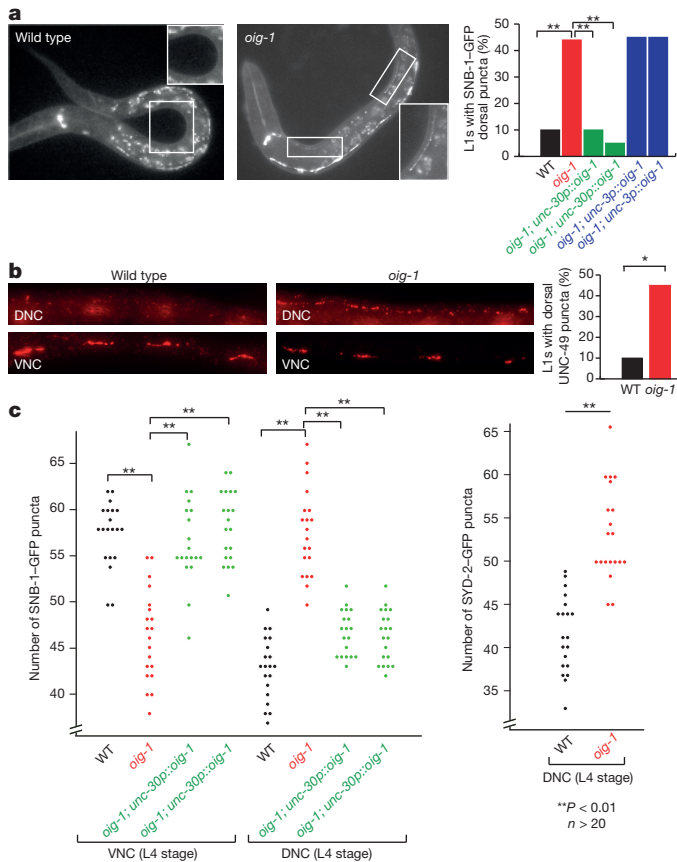


**Figure 2 | Expression of *oig-1* correlates with the inhibition of dorsal synapse formation and is controlled by *unc-30*, *lin-14* and *unc-55*.** **a**, *oig-1* locus, *oig-1* fosmid-based reporters (SL2-NLS-GFP fusion to assess gene regulation; GFP fusion after signal sequence to assess protein localization), deletion allele and *oig-1* fosmid::*sl2::nls::gfp* expression pattern (similar results with 2 independent lines). **b**, *oig-1* expression is regulated by the transcription factors *unc-30* (0/20 animal express the reporter at any stage), *lin-14* (3/20 L1 stage animals express the reporter but at reduced levels; scored in progeny from *lin-14* null mothers carrying a *lin-14* rescue array<sup>23</sup>,  $n = 20$ ), and *unc-55* (0/20 animal express the reporter in VD MNs; 20/20 animals express reporter in DD neurons), but not by *alr-1* (20/20 animal express the reporter), in which VD identity but not synaptic wiring is affected<sup>24</sup>. In *lin-4* animals, *oig-1* expression in DDs persists into the adult stage (20/20 adult animal express the reporter). Original magnification,  $\times 630$ .

conceivable that *unc-55* may also control *oig-1* expression indirectly, through the homeobox gene *irx-1*, a target of *unc-55* (ref. 13). A 125-base pair element that still recapitulates spatiotemporal control in the D-type MN expression contains two sites with partial match to the UNC-30 binding site<sup>14</sup>, and one is required for DD MN expression (Extended Data Fig. 4).

The expression pattern of *oig-1* and its regulation by transcription factors that regulate synapse formation make *oig-1* a candidate for involvement in the synapse-organizing activity of these transcription factors. Animals that carry an *oig-1* deletion allele (Fig. 2a) are viable and fertile, but display locomotory defects and hypersensitivity to the drug aldicarb (Extended Data Fig. 5a, b), which is characteristic of abnormalities in GABAergic signalling<sup>15</sup>. In embryonic and L1 stages when only the DD MNs are present, the presynaptic vesicle proteins SNB-1 and RAB-3 are aberrantly clustered along the dorsal nerve cord of *oig-1* null mutants (Fig. 3a; Extended Data Fig. 2c). Moreover, the postsynaptic GABA receptor UNC-49, which normally clusters on ventral muscle at the L1 stage<sup>16</sup>, clusters ectopically along the dorsal nerve cord (Fig. 3b). Therefore, *oig-1*—like its upstream regulators *unc-30* and *lin-14*—is required to prevent premature DD synapse formation in the dorsal nerve cord.

Examining the synapses of the VD MN that normally exclusively form synapses on ventral muscle, we observed more puncta of three presynaptic markers (SNB-1 and RAB-3 proteins and the Liprin- $\alpha$  protein SYD-2) in the dorsal nerve cord and fewer in the ventral nerve cord of *oig-1* mutants at post-L1 larval stages (Fig. 3c; Extended Data Fig. 2c). This indicates that the VD MNs have aberrant synaptic specializations in the dorsal nerve cord in *oig-1* mutants; although less

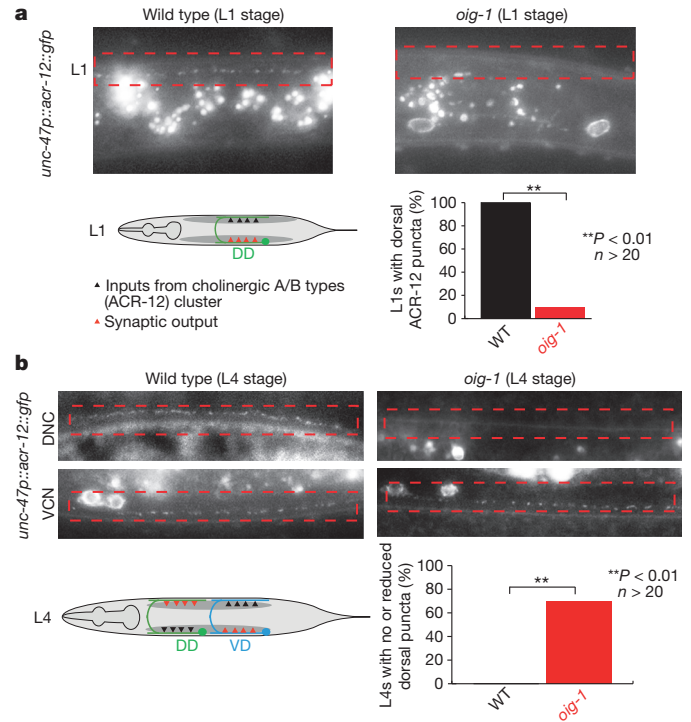


**Figure 3 | Aberrant D-type MN synapse formation in *oig-1* mutants.** **a**, Ectopic DD synapses in the dorsal nerve cord of *oig-1* mutant L1s. **b**, Ectopic UNC-49 localization in *oig-1* mutant L1s, as assessed by UNC-49 antibody staining. UNC-17 staining was used as a control to identify the ventral and dorsal nerve cords of the animals (not shown). **c**, Ectopic VD synapses in the dorsal nerve cord of L4 staged *oig-1* mutants. For each marker in each strain, puncta in the posterior of L4 animals from DD4 to VD11 were scored.  $n > 20$  for each experiment,  $**P < 0.01$ ,  $*P < 0.05$ , original magnification,  $\times 630$ .

severe, this phenotype is similar to those of the VD synaptic defects of *unc-30* and *unc-55* mutants.

We next examined whether *oig-1* function is restricted to controlling the synaptic output of D-type neurons or whether *oig-1* may also affect localization of their synaptic input. Innervation of the DD and VD neurons from the cholinergic A- and B-type neurons can be visualized with GFP-tagged ACR-12 protein<sup>17</sup> which localizes to puncta in the DD neurons in the dorsal nerve cord in the L1 stage, indicative of the cholinergic input from the DA/DB MNs (Fig. 4a). Remarkably, in *oig-1* mutants, these dorsal puncta are not observed (Fig. 4a). In post L1-stage wild-type animals, ACR-12 protein normally labels synapses from DA/DB to the VD neurons in the dorsal cord and synapses from VA/VB to the DD neurons in the ventral cord (Fig. 4b). In *oig-1* mutants, the dorsal, ACR-12(+) synaptic inputs in the VD neurons also do not form properly (Fig. 4b). The coincidence of synaptic input and synaptic output defects in *oig-1* mutants indicates that the localization of synaptic inputs and outputs are coordinated, and that this coordination requires the OIG-1 protein. As expected, the *oig-1* defects in synaptic innervation, as determined by ACR-12 clustering, are mirrored by loss of the temporal (*lin-14*) and spatial (*unc-55*) specificity regulators of *oig-1* expression (Extended Data Fig. 6a, b).

The synaptic defects (as well as the locomotory defects) of *oig-1* mutants can be rescued by expressing *oig-1* specifically in the D-type MN under control of the *unc-30* promoter, whereas expression under control of a cholinergic A- and B-type MN promoter does not rescue

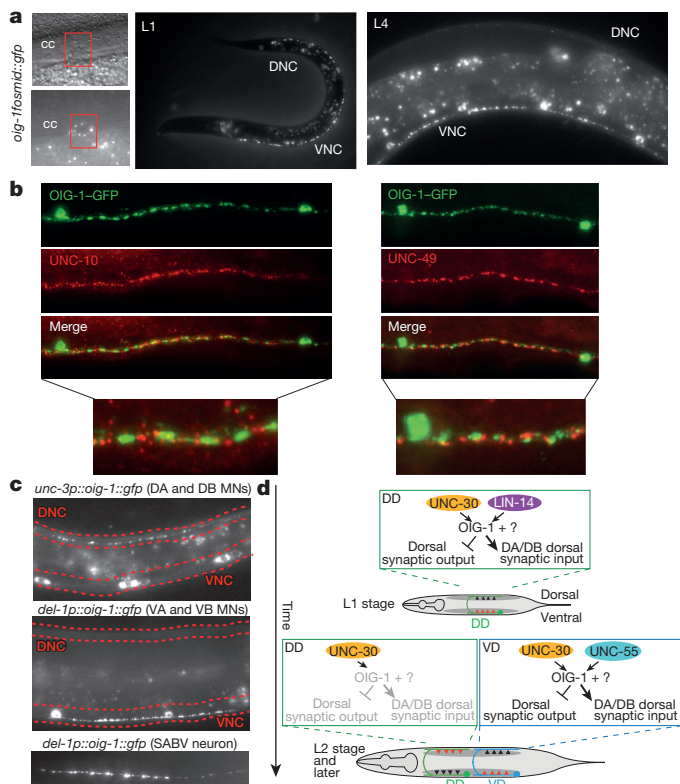


**Figure 4 | OIG-1 also controls cholinergic innervation into the DD and VD neurons.** **a**, **b**, Cholinergic innervation to the D-type MNs is visualized with an *unc-47p::acr-12::gfp* reporter transgene, maintained in an *acr-12(ok367)* mutant background<sup>17</sup>, in L1 stage animals (**a**) and L4 stage animals (**b**).  $n > 20$  for each experiment,  $**P < 0.01$ ,  $*P < 0.05$ , original magnification,  $\times 630$ .

(Fig. 3a, c and Extended Data Fig. 5a, c). Since OIG-1 is predicted to encode a secreted protein, the lack of rescue of *oig-1* with a cholinergic promoter suggests that OIG-1 protein functions cell-autonomously in/on the GABAergic DD and VD classes of MNs and argues against a long-range, diffusible function of OIG-1. Consistent with this autonomy, *oig-1* mutants show no defects in the localization of synapses of the adjacent cholinergic MNs (data not shown). The rescuing activity of *oig-1* critically depends on the integrity of the IgC2 domain (Extended Data Fig. 7). Forced expression of *oig-1* in the D-type MNs under control of two promoters that are not downregulated in the D-type motor neurons (*unc-25* and *unc-30* promoters) is not sufficient to prevent the formation of dorsal synaptic outputs of the DD neurons (data not shown), indicating that *oig-1* collaborates with other factors to regulate synapse formation.

The synaptic wiring defects in *oig-1* mutants suggested that the OIG-1 protein might be localized in a spatially restricted manner in the DD and VD MNs. A fosmid-based reporter in which OIG-1 protein is fused to GFP (and which rescues *oig-1* locomotory defects and the aldicarb hypersensitivity; Fig. 2a and Extended Data Fig. 5a, b) shows punctate localization along the processes of the D-type MNs (Fig. 5a). The punctate pattern of OIG-1 in D-type MNs revealed a surprising localization pattern along the D-type processes. At the L1 stage, OIG-1 is not localized along the dorsal processes (in which the synaptic wiring defects are observed), but is localized along the ventral cord. After the generation of the VD MNs (and extinction of OIG-1 expression in the DD MNs, as described above), OIG-1 protein also localized in the VD neurons along the ventral cord. Co-labelling with the presynaptic RIM protein UNC-10 and the postsynaptic GABA receptor UNC-49 demonstrates that these puncta correspond to the perisynaptic region of synapses that D-type MNs form onto ventral muscle (Fig. 5b). When ectopically expressed in excitatory cholinergic VNC or head MNs, OIG-1-GFP is also targeted to synaptic specializations (Fig. 5c), demonstrating that OIG-1 localization is not dependent





**Figure 5 | OIG-1 localization and summary.** **a**, OIG-1 localization as assessed by an *oig-1* translational reporter fusion protein shown in Fig. 2a. cc, coelomocytes, which take up secreted proteins, including synaptically localized proteins<sup>25</sup>. **b**, OIG-1-GFP localizes perisynaptically as indicated by co-staining with UNC-10 (left panels) and UNC-49 (right panels) antibodies. **c**, OIG-1 is targeted to presynaptic specializations when ectopically expressed in cholinergic neurons. Original magnification,  $\times 630$ . **d**, Summary. Genetic interactions shown here may be indirect and may involve intermediary factors, such as *irx-1* (ref. 13).

on GABAergic-specific synaptic features, but rather contains synaptic targeting properties that are independent of the type of synapse. Taken together, the perisynaptic localization pattern of OIG-1 indicates that a highly localized, synaptic organizer protein is capable of orchestrating the wiring properties of an entire neuron, by promoting synaptic input and preventing ectopic synaptic output in a distal portion of the neuronal process (Fig. 5b). Punctate OIG-1 protein is also observed in a few head neurons (Extended Data Fig. 8).

The mutant phenotype of *oig-1*, specifically the aberrant formation of synaptic output from dorsal DD and VD axons, resembles the mutant phenotype observed upon removal of the SAD-1 kinase<sup>18,19</sup>. In animals lacking *sad-1* or lacking *strd-1*/STRAD $\alpha$ , a pseudokinase required for SAD-1 localization<sup>20</sup>, OIG-1 clusters ectopically along the dorsal nerve cord (Extended Data Fig. 9). Conversely, loss of *oig-1* does not affect localization of SAD-1 (data not shown). What sets OIG-1 apart from these molecules is that, in contrast to the pan-neuronally expressed SAD-1 and STRAD $\alpha$ , OIG-1 seems to operate as a spatio-temporally controlled nexus of this pathway that determines the spatio-temporal specificity of SAD-1 protein function in the context of the D-type MNs. The ability of ventrally and presynaptically localized OIG-1 to organize distally located synaptic inputs and outputs on the dorsal neurite suggests that OIG-1 may trigger a cascade of downstream signalling events or anchor factors on the ventral side which would otherwise contribute to synapse organization on the dorsal side.

In conclusion, we have shown here that three different transcription factors cooperate in an intersectional manner in defined spatial and temporal contexts to control the expression of a perisynaptically localized organizer molecule, OIG-1, which orchestrates the localization of

synaptic outputs and inputs of two different neuron types (Fig. 5d). *unc-30* needs to cooperate with other transcription factors and these collaborators confer spatiotemporal specificity. In embryonic and L1 stages, spatially but not temporally restricted *unc-30* cooperates with temporally, but not spatially controlled, *lin-14* to prevent DD MN synapse assembly at the inappropriate location via induction of *oig-1* expression. After the L1 stage, DD/VD-expressed *unc-30* collaborates with the subtype (VD)-specific *unc-55* transcription factor to restrict *oig-1* expression to the VD neurons where it organizes synaptic inputs and outputs (Fig. 5d).

Our findings demonstrate that the localization of synaptic inputs and outputs of a neuron are coordinated and that this coordination is apparently achieved, at least in part, by the OIG-1 organizer protein. It will be interesting to examine whether similar synaptic organizer functions can be ascribed to any of the multiple small, single immunoglobulin domain proteins, some secreted, some transmembrane, which are encoded in the *C. elegans*<sup>21</sup> and vertebrate genomes<sup>22</sup>.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 17 September 2014; accepted 12 May 2015.

Published online 17 June 2015.

- White, J. G., Albertson, D. G. & Anness, M. A. Connectivity changes in a class of motoneuron during the development of a nematode. *Nature* **271**, 764–766 (1978).
- Eastman, C., Horvitz, H. R. & Jin, Y. Coordinated transcriptional regulation of the unc-25 glutamic acid decarboxylase and the unc-47 GABA vesicular transporter by the *Caenorhabditis elegans* UNC-30 homeodomain protein. *J. Neurosci.* **19**, 6225–6234 (1999).
- Jin, Y., Hoskins, R. & Horvitz, H. R. Control of type-D GABAergic neuron differentiation by *C. elegans* UNC-30 homeodomain protein. *Nature* **372**, 780–783 (1994).
- Hallam, S. J. & Jin, Y. *lin-14* regulates the timing of synaptic remodelling in *Caenorhabditis elegans*. *Nature* **395**, 78–82 (1998).
- Ruvkun, G. & Giusto, J. The *Caenorhabditis elegans* heterochronic gene *lin-14* encodes a nuclear protein that forms a temporal developmental switch. *Nature* **338**, 313–319 (1989).
- Walthall, W. W. & Plunkett, J. A. Genetic transformation of the synaptic pattern of a motoneuron class in *Caenorhabditis elegans*. *J. Neurosci.* **15**, 1035–1043 (1995).
- Hedgecock, E. M., Culotti, J. G., Hall, D. H. & Stern, B. D. Genetics of cell and axon migrations in *Caenorhabditis elegans*. *Development* **100**, 365–382 (1987).
- Aurelio, O., Hall, D. H. & Hobert, O. Immunoglobulin-domain proteins required for maintenance of ventral nerve cord organization. *Science* **295**, 686–690 (2002).
- Wightman, B., Ha, I. & Ruvkun, G. Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell* **75**, 855–862 (1993).
- Zhou, H. M. & Walthall, W. W. UNC-55, an orphan nuclear hormone receptor, orchestrates synaptic specificity among two classes of motor neurons in *Caenorhabditis elegans*. *J. Neurosci.* **18**, 10438–10444 (1998).
- Shan, G., Kim, K., Li, C. & Walthall, W. W. Convergent genetic programs regulate similarities and differences between related motor neuron classes in *Caenorhabditis elegans*. *Dev. Biol.* **280**, 494–503 (2005).
- Araya, C. L. *et al.* Regulatory analysis of the *C. elegans* genome with spatiotemporal resolution. *Nature* **512**, 400–405 (2014).
- Petersen, S. C. *et al.* A transcriptional program promotes remodeling of GABAergic synapses in *Caenorhabditis elegans*. *J. Neurosci.* **31**, 15362–15375 (2011).
- Cinar, H., Keles, S. & Jin, Y. Expression profiling of GABAergic motor neurons in *Caenorhabditis elegans*. *Curr. Biol.* **15**, 340–346 (2005).
- Miller, K. G. *et al.* A genetic selection for *Caenorhabditis elegans* synaptic transmission mutants. *Proc. Natl Acad. Sci. USA* **93**, 12593–12598 (1996).
- Bamber, B. A., Beg, A. A., Twyman, R. E. & Jorgensen, E. M. The *Caenorhabditis elegans* unc-49 locus encodes multiple subunits of a heteromultimeric GABA receptor. *J. Neurosci.* **19**, 5348–5359 (1999).
- Petrash, H. A., Philbrook, A., Haburcak, M., Barbaggio, B. & Francis, M. M. ACR-12 ionotropic acetylcholine receptor complexes regulate inhibitory motor neuron activity in *Caenorhabditis elegans*. *J. Neurosci.* **33**, 5524–5532 (2013).
- Crump, J. G., Zhen, M., Jin, Y. & Bargmann, C. I. The SAD-1 kinase regulates presynaptic vesicle clustering and axon termination. *Neuron* **29**, 115–129 (2001).
- Kim, J. S., Hung, W. & Zhen, M. The long and the short of SAD-1 kinase. *Commun. Integr. Biol.* **3**, 251–255 (2010).
- Kim, J. S., Hung, W., Narbonne, P., Roy, R. & Zhen, M. *C. elegans* STRAD $\alpha$  and SAD cooperatively regulate neuronal polarity and synaptic organization. *Development* **137**, 93–102 (2010).
- Hobert, O. The neuronal genome of *Caenorhabditis elegans*. *WormBook*. <http://dx.doi.org/10.1895/wormbook.1.161.1> (2013).
- Letunic, I., Doerks, T. & Bork, P. SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res.* **40**, D302–D305 (2012).

23. Hong, Y., Lee, R. C. & Ambros, V. Structure and function analysis of LIN-14, a temporal regulator of postembryonic developmental events in *Caenorhabditis elegans*. *Mol. Cell. Biol.* **20**, 2285–2295 (2000).
24. Melkman, T. & Sengupta, P. Regulation of chemosensory and GABAergic motor neuron development by the *C. elegans* *Aristaless/Arx* homolog *alr-1*. *Development* **132**, 1935–1949 (2005).
25. Rapti, G., Richmond, J. & Bessereau, J. L. A single immunoglobulin-domain protein required for clustering acetylcholine receptors in *C. elegans*. *EMBO J.* **30**, 706–718 (2011).

**Acknowledgements** We thank the TransgeneOme project, the CGC, J. L. Bessereau, M. Gendrel, S. Y. Kerk for reagents, Q. Chen for microinjection, E. Southgate and N. Thomson for electron microscopy analyses, S. Brenner for first noting the neural defects of *unc-30(e191)*, E. Yemini for assistance with worm tracking, and D. Miller,

S. He, I. Greenwald and members of the Hobert laboratory for comments on this manuscript. This work was funded by the National Institutes of Health (R01NS039996-05 and R01NS050266-03), the Howard Hughes Medical Institute and the UK Medical Research Council.

**Author Contributions** K.H. conducted all the experiments with the exception of the electron microscopy analysis, done by J.G.W.; O.H. supervised the study and all authors contributed to writing of the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.H. (kh2512@columbia.edu) or O.H. (or38@columbia.edu).

## METHODS

**C. elegans strains.** Worms were grown at 20 °C on nematode growth media (NGM) plates seeded with bacteria (*Escherichia coli* OP50) as a food source. L1 animals were obtained by hypochlorite-treating gravid adult animals and letting embryos hatch and arrest in M9 for 16–18 h.

Mutant alleles used in this study: LGI: *unc-55(e1170)*, LGII: *lin-4(e912)*, LGIII: *oig-1(ok1687)*, *strd-1(ok2283)*, LGIV: *unc-30(191)*, LGX: *lin-14(ma135)*, *alr-1(oy42)*, *sad-1(ky289)*, *acr-12(ok367)*

**Transgenes.** *otEx5663 (unc-3<sup>2.4kbprom</sup>::gfp::rab-3::unc-10<sup>3'UTR</sup>)* used in Fig. 1, Extended Data Fig. 2, *otEx4816 (oig-1<sup>fsmid</sup>::sl2::gfp)*, *otEx450 (oig-1<sup>fsmid</sup>::sl2::gfp)* used in Fig. 2, *otEx5651 (unc-30<sup>2.4kbprom</sup>::gfp)* used in Extended Data Fig. 3a, *wgl395 (unc-30<sup>fsmid</sup>::TY1::gfp::3xFLAG)* used in Extended Data Fig. 3a, *otEx5765 (lin-14<sup>fsmid</sup>::gfp)* used in Extended Data Fig. 3b, *juIs1 (unc-25p::snb-1::gfp)*<sup>4</sup> used in Fig. 3a, c, Extended Data Figs 5c and 7, *otEx4955 (unc-30<sup>2.4kbprom</sup>::oig-1 line 1)* used in Fig. 3a, c and Extended Data Fig. 5a, c, *otEx4956 (unc-30<sup>2.4kbprom</sup>::oig-1 line 2)* used in Fig. 3a, c, *otEx4941 (unc-3<sup>prom</sup>::oig-1 line 1)* used in Fig. 3a, c and Extended Data Fig. 5c, *otEx4942 (unc-3<sup>prom</sup>::oig-1 line 2)* used in Fig. 3a, c, *hpl3 (unc-25p::syd-2::gfp)*<sup>26</sup> used in Fig. 3c, *ufls92 (unc-47<sup>prom</sup>::acr-12::gfp)*<sup>17</sup> used in Fig. 4 and Extended Data Fig. 6, *otEx5664 (oig-1<sup>fsmid</sup>::gfp)* used in Fig. 5a, b and Extended Data Figs 5a, b, 8 and 9, *otEx5858 (unc-3<sup>588bprom</sup>::OIG-1::gfp)* used in Fig. 5c, *otEx5859 (del-1<sup>488bprom</sup>::oig-1::gfp)* used in Fig. 5c, *otEx6212 (unc-30<sup>2.4kbprom</sup>::oig-1E64A line 1)*, *otEx6213 (unc-30<sup>2.4kbprom</sup>::oig-1E64A line 2)*, *otEx6214 (unc-30<sup>2.4kbprom</sup>::oig-1W75A line 1)* and *otEx6215 (unc-30<sup>2.4kbprom</sup>::oig-1W75A line 2)* all used in Extended Data Fig. 7.

*oig-1 cis-regulatory analysis constructs* (all used in Extended Data Fig. 4): *otEx5993 (oig-1<sup>prom1</sup>::NLS::gfp line 1)*, *otEx5994 (oig-1<sup>prom1</sup>::NLS::gfp line 2)*, *otEx5995 (oig-1<sup>prom1</sup>::NLS::gfp line 3)*, *otEx5996 (oig-1<sup>prom2</sup>::NLS::gfp line 1)*, *otEx5997 (oig-1<sup>prom2</sup>::NLS::gfp line 2)*, *otEx5998 (oig-1<sup>prom2</sup>::NLS::gfp line 3)*, *otEx6003 (oig-1<sup>prom2</sup>::NLS::gfp line 1)*, *otEx6004 (oig-1<sup>prom3</sup>::NLS::gfp line 2)*, *otEx6005 (oig-1<sup>prom3</sup>::NLS::gfp line 3)*, *otEx6006 (oig-1<sup>prom4</sup>::NLS::gfp line 1)*, *otEx6007 (oig-1<sup>prom4</sup>::NLS::gfp line 2)*, *otEx6008 (oig-1<sup>prom4</sup>::NLS::gfp line 3)*, *otEx6009 (oig-1<sup>prom5</sup>::NLS::gfp line 1)*, *otEx6010 (oig-1<sup>prom5</sup>::NLS::gfp line 2)*, *otEx6011 (oig-1<sup>prom5</sup>::NLS::gfp line 3)*, *otEx6034 (oig-1<sup>prom6</sup>::NLS::gfp line 1)*, *otEx6035 (oig-1<sup>prom6</sup>::NLS::gfp line 2)*, *otEx6036 (oig-1<sup>prom6</sup>::NLS::gfp line 3)*, *otEx6037 (oig-1<sup>prom7</sup>::NLS::gfp line 1)*, *otEx6038 (oig-1<sup>prom7</sup>::NLS::gfp line 2)*, *otEx6039 (oig-1<sup>prom7</sup>::NLS::gfp line 3)*, *otEx6060 (oig-1<sup>prom8</sup>::NLS::gfp line 1)*, *otEx6061 (oig-1<sup>prom8</sup>::NLS::gfp line 2)*, *otEx6121 (oig-1<sup>prom9</sup>::NLS::gfp line 1)*, *otEx6122 (oig-1<sup>prom9</sup>::NLS::gfp line 2)*, *otEx6079 (oig-1<sup>prom10</sup>::NLS::gfp line 1)*, *otEx6080 (oig-1<sup>prom10</sup>::NLS::gfp line 2)*, *otEx6081 (oig-1<sup>prom10</sup>::NLS::gfp line 3)*.

**Generation of *oig-1* transgenes.** The *oig-1<sup>fsmid</sup>::sl2::gfp* reporter (shown in Fig. 2) was generated by fosmid recombineering using the fosmid WRM0614C07 and a SL2-based, nuclear-localized *gfp* cassette, pBALU9<sup>27</sup>. The inclusion of the SL2 sequence results in the production of nuclear localized GFP. The reporter was injected at 10 ng  $\mu\text{l}^{-1}$  with *rol-6(su1006)* at 2 ng  $\mu\text{l}^{-1}$  and sonicated OP50 genomic DNA at 120 ng  $\mu\text{l}^{-1}$ . An extrachromosomal array (*otEx4816*) was integrated to yield *otIs450 IV*.

The *oig-1* translational fosmid *gfp* reporter, *oig-1<sup>fsmid</sup>::gfp* (shown in Fig. 5) was generated by fosmid recombineering using the fosmid WRM0614C07 and a *gfp* cassette pBALU25, modified from pBALU1<sup>27</sup>. pBALU25 was created by mutating the coding sequence of *gfp* in pBALU1 to contain the amino acid changes F64L and S65T. This cassette was recombineered and inserted into the *oig-1* fosmid immediately following the predicted signal peptide sequence (after the 72nd base pair of *oig-1*). This translational reporter was injected at 10 ng  $\mu\text{l}^{-1}$  with *ttx-3<sup>prom</sup>::mCherry* at 3 ng  $\mu\text{l}^{-1}$  and sonicated OP50 genomic DNA at 120 ng  $\mu\text{l}^{-1}$ .

The *unc-3<sup>558bprom</sup>::oig-1::gfp* construct was generated by TOPO cloning a PCR fusion of 558 bp upstream of the *unc-3* translational start site with a fragment of the *oig-1<sup>fsmid</sup>::gfp* construct, containing from the *oig-1* translational start site to 500 bp downstream of the stop codon. This construct was PCR-amplified from the start of the *unc-3* promoter to 500 bp downstream of the *oig-1* locus and injected at 10 ng  $\mu\text{l}^{-1}$  with *ttx-3<sup>prom</sup>::mCherry* at 3 ng  $\mu\text{l}^{-1}$  and sonicated OP50 genomic DNA at 120 ng  $\mu\text{l}^{-1}$ .

The *del-1<sup>488bprom</sup>::oig-1::gfp* construct was generated by TOPO cloning as the *unc-3<sup>558bprom</sup>::oig-1::gfp* with 488 bp upstream of the *del-1* translational start site. This construct was PCR-amplified from the start of the *del-1* promoter to 500 bp downstream of the *oig-1* locus and injected at 10 ng  $\mu\text{l}^{-1}$  with *ttx-3<sup>prom</sup>::mCherry* at 3 ng  $\mu\text{l}^{-1}$  and sonicated OP50 genomic DNA at 120 ng  $\mu\text{l}^{-1}$ .

The *unc-30p::oig-1* construct was generated by cloning the 2.4 kb *unc-30* promoter into the EcoRV site of the first MCS of pPD49.26 and the *oig-1* locus from the start to stop codon (1,521 bp) into the BamHI site of the second MCS. This construct was digested with PvuII and injected at 5 ng  $\mu\text{l}^{-1}$  with *myo-2p::gfp* at 3 ng  $\mu\text{l}^{-1}$  and OP50 genomic DNA at 120 ng  $\mu\text{l}^{-1}$ . Site-directed mutagenesis of this construct was used to generate *unc-30p::oig-1E64A* and *unc-30p::oig-1W75A*. These constructs were injected as described above.

The *unc-3p::oig-1* construct was generated by cloning the 558 bp *unc-3* promoter into the EcoRV site of the first MCS of pPD49.26 and the *oig-1* locus from the start to stop codon (1,521 bp) into the BamHI site of the second MCS. This construct was digested with PvuII and injected at 10 ng  $\mu\text{l}^{-1}$  with *myo-2p::gfp* at 3 ng  $\mu\text{l}^{-1}$  and OP50 genomic DNA at 120 ng  $\mu\text{l}^{-1}$ .

The *oig-1* promoter deletion constructs were generated by cloning the various promoter fragments into the HindIII and BamHI sites of the MSC of a 2XNLSGFP plasmid. Promoter constructs with potential UNC-30 binding sites deleted were generated using site-directed mutagenesis. These constructs were injected at 50 ng  $\mu\text{l}^{-1}$  with *rol-6* at 30 ng  $\mu\text{l}^{-1}$  and *rol-6(su1006)* at 20 ng  $\mu\text{l}^{-1}$ .

**Wormtracker assays.** Tracking assays were performed as previously described<sup>28</sup>. Briefly, L4 animals were placed on an NGM plate seeded with 20  $\mu\text{l}$  of OP50 bacteria in the centre. Automated tracking was performed with the Worm Tracker 2.0 (WT2) which uses a camera to track and record individual worms. Twenty worms of each genotype were tracked for 5 min each at 20 °C. Analysis was performed as previously described<sup>28</sup>.

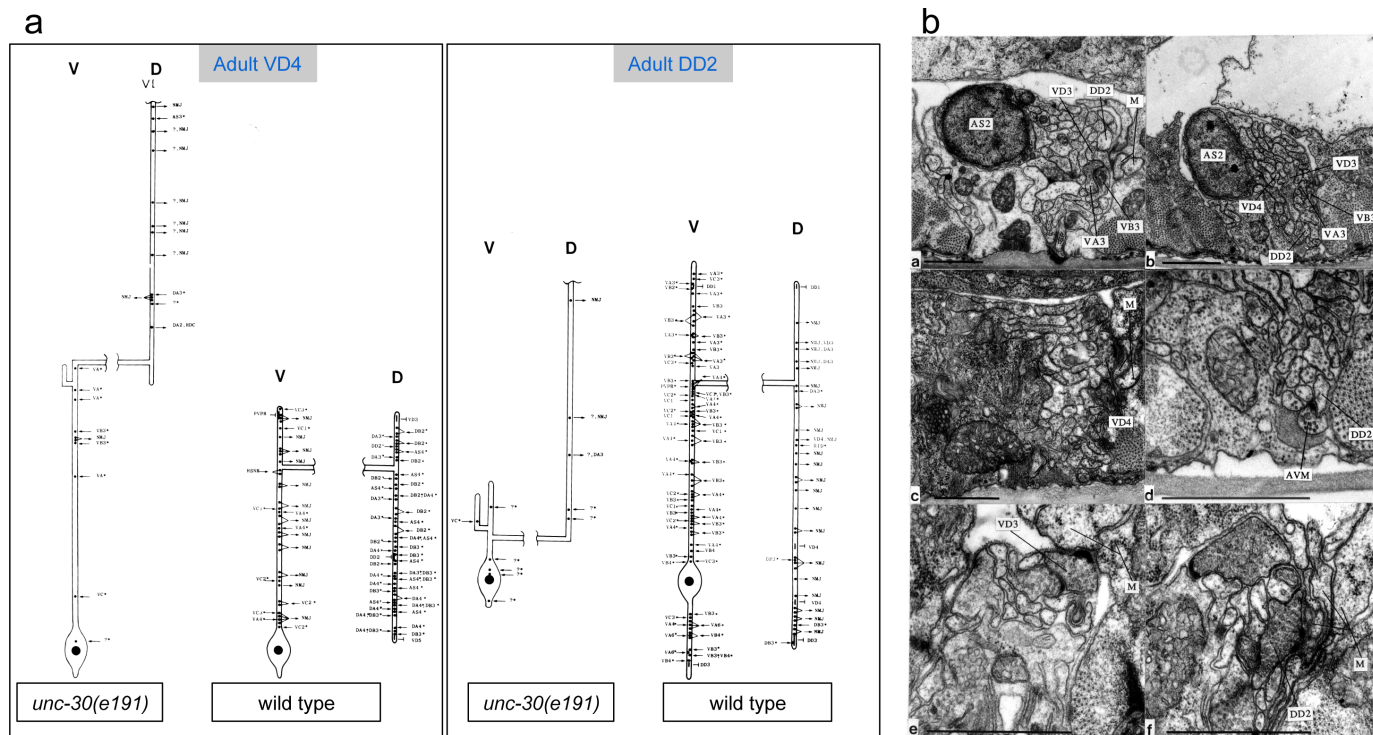
**Aldicarb assays.** Aldicarb assays were performed as previously described<sup>33</sup>. Briefly, ~20 young adult animals (24 h after L4 stage, blinded for genotype) were picked for freshly seeded NGM plates containing 1 mM aldicarb (ChemService). Worms were assayed for paralysis every 15 min by prodding with a platinum wire. A worm was considered paralyzed if it did not respond to prodding to the head and tail three times each at a given time point. Strains were grown and assayed at 20 °C.

**Antibody staining.** Antibody staining was performed as previously described<sup>29</sup>. Briefly, following a freeze-crack procedure, worms were fixed by a treatment in ice-cold acetone for 5 min and then ice-cold methanol for 5 min. Worms were collected in 1× PBS and centrifuged briefly. The PBS was removed and worms were incubated in a blocking solution (1× PBS, 0.2% gelatin, 0.25% Triton X-100) for 30 min at 20 °C. After the blocking solution was removed, worms were incubated with primary antibodies diluted in PGT (1× PBS, 0.1% gelatin, 0.25% Triton X-100) overnight at 4 °C. The anti-UNC-49 antibody<sup>34</sup> was used at a 1:500 dilution. The anti-UNC-17 antibody<sup>30</sup> was used at a 1:3,000 dilution. The anti-GFP antibody (Life Technologies A10262) was used at 1:1,000. The anti-RIM2 (used to recognize UNC-10) was used at a 1:10 dilution (Developmental Studies Hybridoma Bank – University of Iowa). Worms were washed 5 times in wash solution (1× PBS, 0.25% Triton X-100) for 20 min each wash. Worms were then incubated with secondary antibodies diluted 1:1,000 in PGT for 3 h at 20 °C. Alexa Fluor 488 goat anti-chicken (Invitrogen A11039) was used to detect the anti-GFP antibody. Alexa Fluor 594 donkey anti-mouse (Invitrogen A-21203) was used to detect the anti-RIM2 antibody. Alexa Fluor 555 donkey anti-rabbit (Invitrogen A-31572) was used to detect the anti-UNC-49 antibody. Alexa Fluor 488 donkey anti-mouse (Invitrogen A-21202) was used to detect the anti-UNC-17 antibody. Worms were then washed 5 times for 20 min each wash. Following the final wash, worms were mounted in Fluorogel with Tris buffer (Electron Microscopy Sciences).

**Statistical analysis.** For results shown in Figs 3a, b, 4a, b, 6a, c, 7 and 9, we performed Fisher's exact test,  $**P < 0.01$ ,  $*P < 0.05$ . For results shown in Fig. 3c, Extended Data Fig. 2c, we performed a Student's *t*-test (2 sided, type 2),  $**P < 0.01$ ,  $*P < 0.05$ . For WormTracker analysis in Extended Data Fig. 3a, we used Wilcoxon rank sum to test the differences between *oig-1*, wild-type, and rescued strains,  $**q < 0.01$ ,  $*q < 0.05$ . No statistical methods were used to predetermine sample size, and the experiments were not randomized.

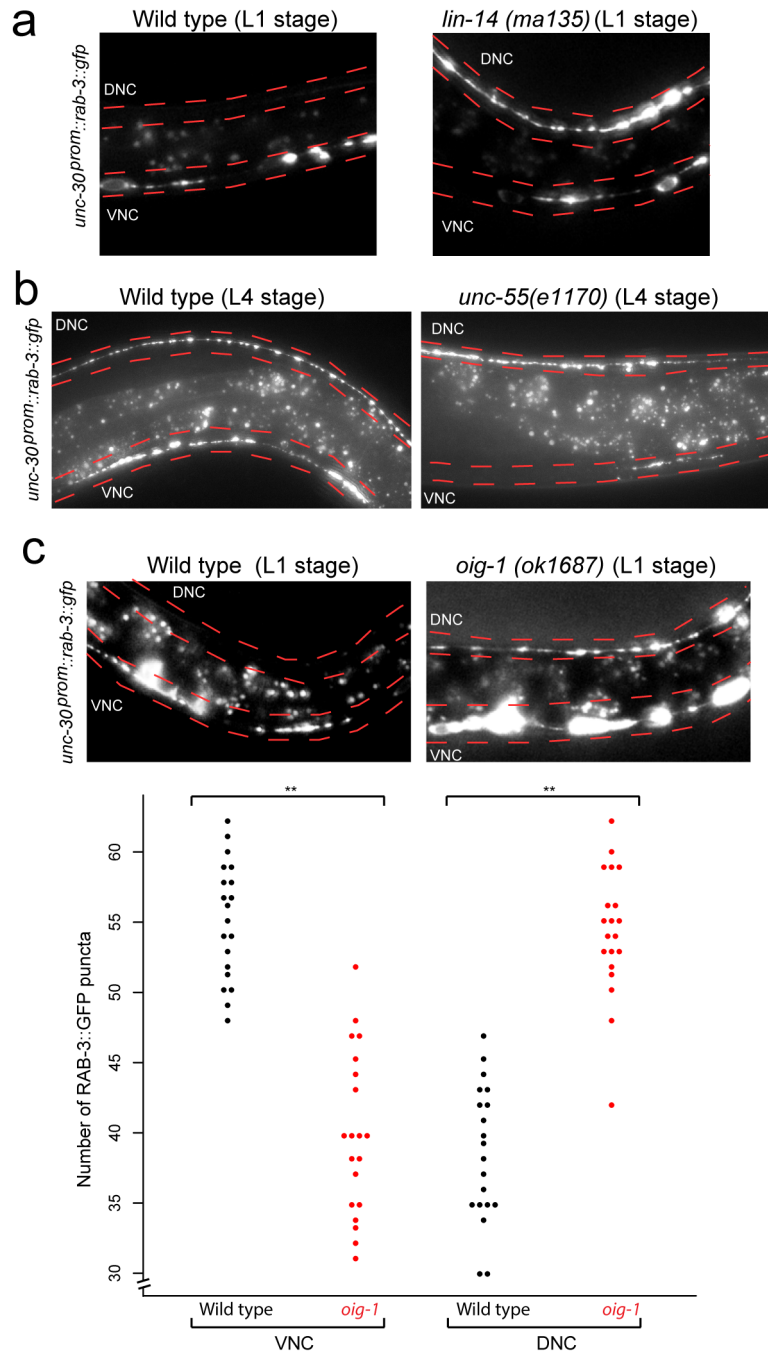
- Yeh, E., Kawano, T., Weimer, R. M., Bessereau, J. L. & Zhen, M. Identification of genes involved in synaptogenesis using a fluorescent active zone marker in *Caenorhabditis elegans*. *J. Neurosci.* **25**, 3833–3841 (2005).
- Tursun, B., Cochella, L., Carrera, I. & Hobert, O. A toolkit and robust pipeline for the generation of fosmid-based reporter genes in *C. elegans*. *PLoS ONE* **4**, e4625 (2009).
- Yemini, E., Jucikas, T., Grundy, L. J., Brown, A. E. & Schafer, W. R. A database of *Caenorhabditis elegans* behavioral phenotypes. *Nature Methods* **10**, 877–879 (2013).
- Gendrel, M., Rapti, G., Richmond, J. E. & Bessereau, J. L. A secreted complement-control-related protein ensures acetylcholine receptor clustering. *Nature* **461**, 992–996 (2009).
- Duerr, J. S., Han, H. P., Fields, S. D. & Rand, J. B. Identification of major classes of cholinergic neurons in the nematode *Caenorhabditis elegans*. *J. Comp. Neurol.* **506**, 398–408 (2008).
- White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Phil. Trans. R. Soc. Lond. B* **314**, 1–340 (1986).
- Li, J. & Greenwald, I. LIN-14 inhibition of LIN-12 contributes to precision and timing of *C. elegans* vulval fate patterning. *Curr. Biol.* **20**, 1875–1879 (2010).
- Mahoney, T. R., Luo, S. & Nonet, M. L. Analysis of synaptic transmission in *Caenorhabditis elegans* using an aldicarb-sensitivity assay. *Nature Protocols* **1**, 1772–1777 (2006).
- Gally, C. & Bessereau, J. L. GABA is dispensable for the formation of junctional GABA receptor clusters in *Caenorhabditis elegans*. *J. Neurosci.* **23**, 2591–2599 (2003).





**Extended Data Figure 1 | Electron microscopical analysis of *unc-30(e191)* mutants.** **a**, Reconstructions of a VD4 and a DD2 MN from an *unc-30(e191)* animal compared to the same neurons in a wild-type animal. Cell bodies (large black dots) are all situated in the ventral cord. Processes emanate anteriorly (upwards on plots) from the cell body and run along the ventral cord. Lateral branches leave the ventral cord (V) and run round to the dorsal cord (D) as a circumferential commissure (broken horizontal process in the plots). Commissures from *unc-30* mutant type-D neurons are situated in the same regions as those of their wild-type counterparts. However, the cell bodies of DD neurons are often displaced anteriorly in the mutants, with the consequence that DD neurons have shorter processes in the ventral cord (C). Processes in the dorsal cord run anteriorly in mutant animals, whereas they branch with the main branch running posteriorly in wild-type animals. Neuromuscular junctions (NMJs) in *unc-30* mutants are made predominantly in the dorsal cord by both the DDs and VDs, whereas in the wild type, only DD neurons innervate dorsal muscles. The synaptic inputs to the DD and VD neurons in mutant animals (inward pointing arrows for chemical synapses and 'T' for gap junctions) are generally abnormal. The reconstructed DD2 neuron received synapses from several unidentified processes on the ventral side (depicted with a '?'). These processes do not belong to VA or VB neurons, the normal pre-synaptic partners of DD, as all the local VA and VB neurons were identified. From the location and synaptic behaviour of these processes, it is probable that they belong to interneurons which span the length of the cord and do not usually innervate D-type neurons. An asterisk (\*) indicates that a synapse has

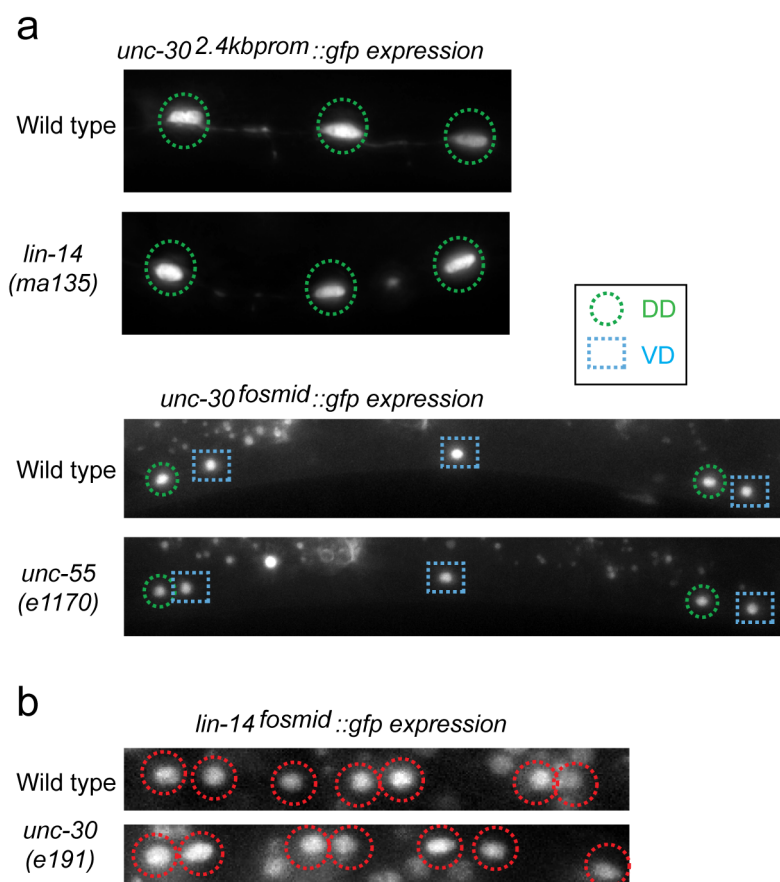
multiple post-synaptic elements. A total of six VDs and three DDs were reconstructed. Each reconstruction covered around 2,000 electron microscopy sections, corresponding to a length of 100  $\mu\text{m}$  along the body of the animal. The *unc-30(e191)* mutation does not affect MN cell body position or the synaptic behaviour of the DA/DB neurons, except in regard to their synapses to D-type neurones; this made it possible to unambiguously identify D type neurones from their positions and by eliminating other identified classes of MNs. Electron microscopy and reconstructions of micrographs of serial sections were performed as described in ref. 31. **b**, Electron micrographs. The processes of DD and VD neurons normally run subjacent to the bounding basal lamina of the ventral cord immediately dorsal to the axons of the VA and VB neurons (**a**). NMJs are made through the basal lamina onto muscle arms (M). In *unc-30(e191)* animals, the axons of the DD and VD neurons wander round the cord and do not run in defined locations; the configuration shown in **b** is typical but not stereotyped. Very few NMJs are made in the ventral cord by the DD or VD neurons in *unc-30* mutants; those that are made look rather small (**c**). Atypical synapses (**d**) are often made onto DD or VD processes from neurons such as the touch receptor neuron AVM (6). It is probable that these synapses are not normally found as the processes of AVM, and DD or VD do not normally run alongside each other in wild-type animals. Both the DD and VD MNs make NMJs to dorsal muscles in the dorsal cord of *unc-30* mutants, whereas in wild-type animals, only DD neurons do so and VD neurons innervate ventral muscles in the ventral cord. **e**, **f**, NMJs made by VD (**e**) and DD (**f**) neurons in the dorsal cord of an *unc-30(e191)* mutant. Scale bars, 1  $\mu\text{m}$ .



**Extended Data Figure 2 | RAB-3 is ectopically localized in *lin-14*, *unc-55*, and *oig-1* mutants.** **a**, Presynaptic marker RAB-3 ectopically localizes to the dorsal nerve cord (DNC; marked in red) in *lin-14* mutant animals. RAB-3-GFP puncta (from *otEx5663, unc-30p::gfp::rab-3*) localize mostly to the ventral nerve cord (VNC; marked in red) in wild-type L1 animals (left). Ectopic RAB-3-GFP puncta localize mostly to the dorsal nerve cord in 95% of *lin-14* L1 animals (right, scored in progeny from *lin-14* null animals carrying a *lin-14* rescue array<sup>23</sup>). Ventral and dorsal nerve cords are indicated by red dotted lines. L1 animals were obtained by hypochlorite-treating gravid adult animals and letting embryos hatch and arrest in M9 for 16–18 h.  $n > 20$  for each strain scored. **b**, RAB-3 ectopically localizes to the dorsal nerve cord in *unc-55* L4 mutant animals. RAB-3-GFP puncta localizes to both the ventral (VNC) and dorsal (DNC) nerve cord in 100% of wild-type L4 animals (left). RAB-3-GFP puncta localize mostly to the dorsal nerve cord in 100% of *unc-55* L4

animals (right). Ventral and dorsal nerve cords are indicated by red dotted lines. Signals between the nerve cords are autofluorescence from the gut.  $n > 20$  for each strain scored. **c**, RAB-3 ectopically localizes to the dorsal nerve cord in *oig-1* mutants. RAB-3 normally localizes to the ventral nerve cord (VNC, marked in red) in wild-type L1 animals (top left). Ectopic RAB-3-GFP puncta localize to the dorsal nerve cord in 55% of *oig-1* L1 animals (top right, compared to 20% of wild-type animals). L1 animals were obtained by hypochlorite-treating gravid adult animals and letting embryos hatch and arrest in M9 for 16–18 h.  $n > 20$  for each strain scored. In wild-type L4 animals, more RAB-3-GFP puncta are localized in the VNC than in the DNC of the animal (bottom, black dots). Conversely, in *oig-1* mutants, more RAB-3-GFP puncta are localized in the DNC than in the VNC (bottom, red dots).

\*\* $P < 0.01$ ,  $n = 20$  for each strain, original magnification,  $\times 630$ .

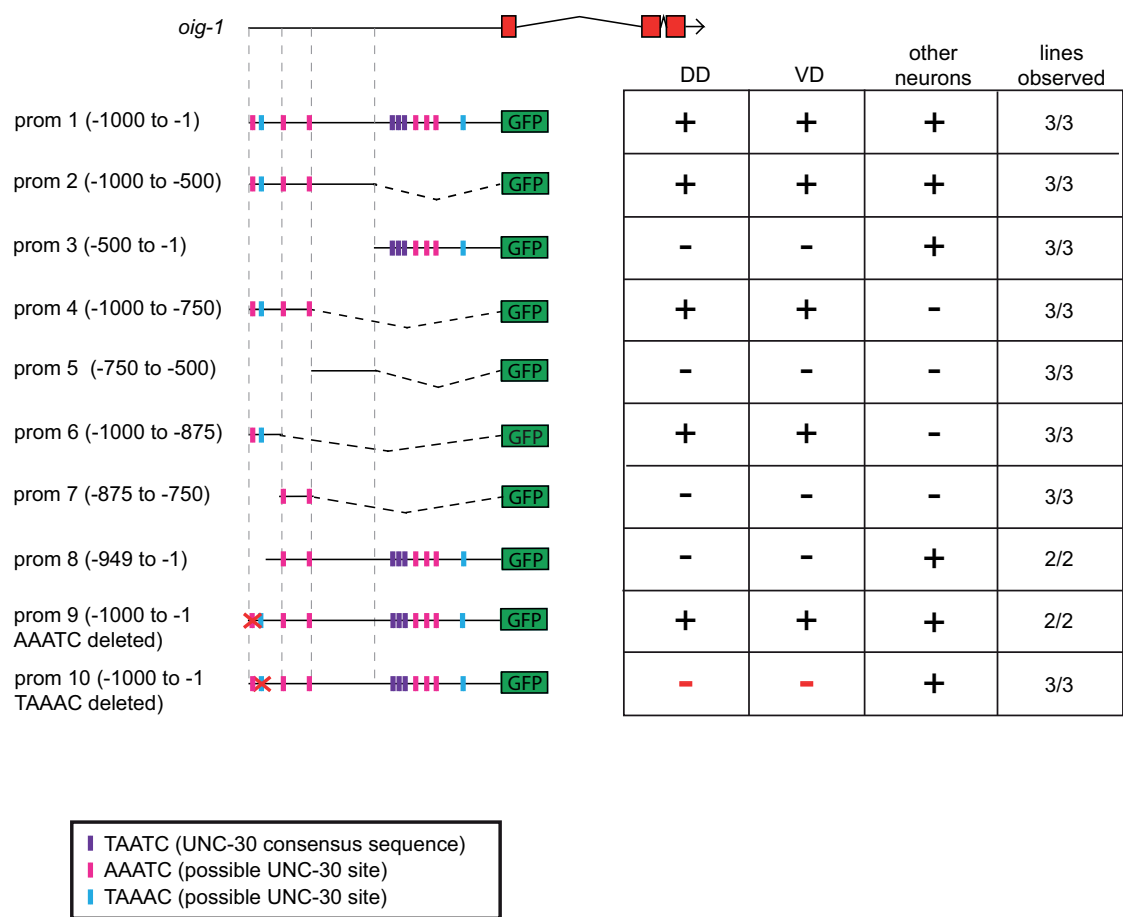


**Extended Data Figure 3 | Mutual independence of transcription factor activities.** **a**, Expression of *unc-30* is not affected by loss of *lin-14* or *unc-55*. A 2.4 kb *unc-30* promoter *gfp* fusion reporter is expressed in the DD MNs (green circles) in wild-type L1 animals; this expression is not affected in *lin-14*(<sup>-</sup>) mutant animals (scored in progeny from *lin-14* null mothers carrying a *lin-14* rescue array<sup>23</sup>). An *unc-30* fosmid-based reporter, kindly provided by the TransgeneOme project, is expressed in the DD (green circles) and VD (blue squares) MNs (DD4 to VD10 shown) in wild-type L4 animals; this expression

is not affected in *unc-55*(*e1170*) L4 animals.  $n > 20$  for each genotype.

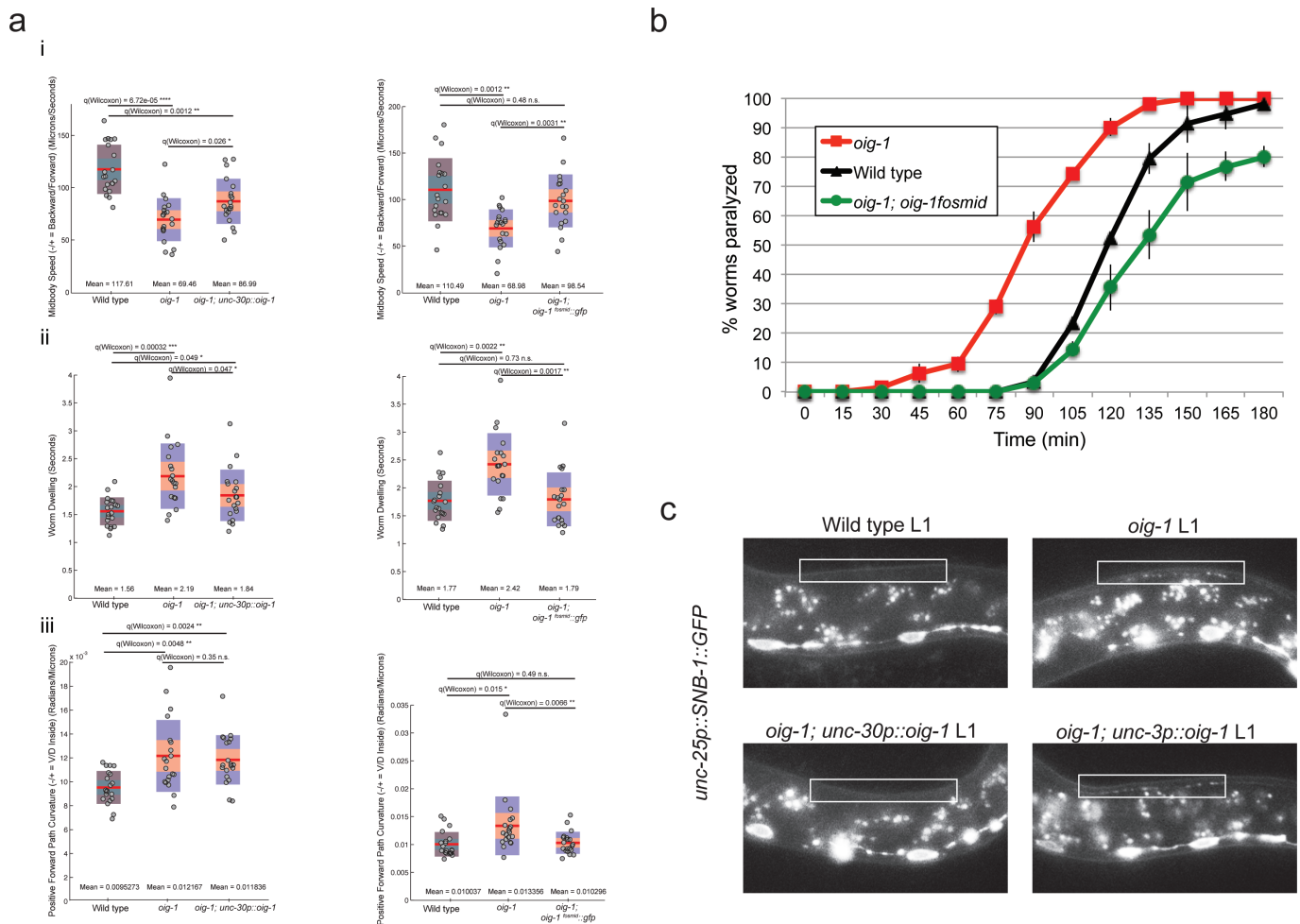
**b**, Expression of a *lin-14* fosmid-based reporter construct<sup>32</sup> is unaffected by loss of *unc-30*. *lin-14* is expressed in the DA, DB, and DD MNs in the VNC at the L1 stage (average number of VNC cells = 15); this expression is not affected in *unc-30*(*e191*) mutant L1s (average number of VNC cells = 15);  $n > 20$  for each genotype. Loss of *unc-30* also does not affect *unc-55* expression, as shown by ref. 11. Original magnification,  $\times 630$ .





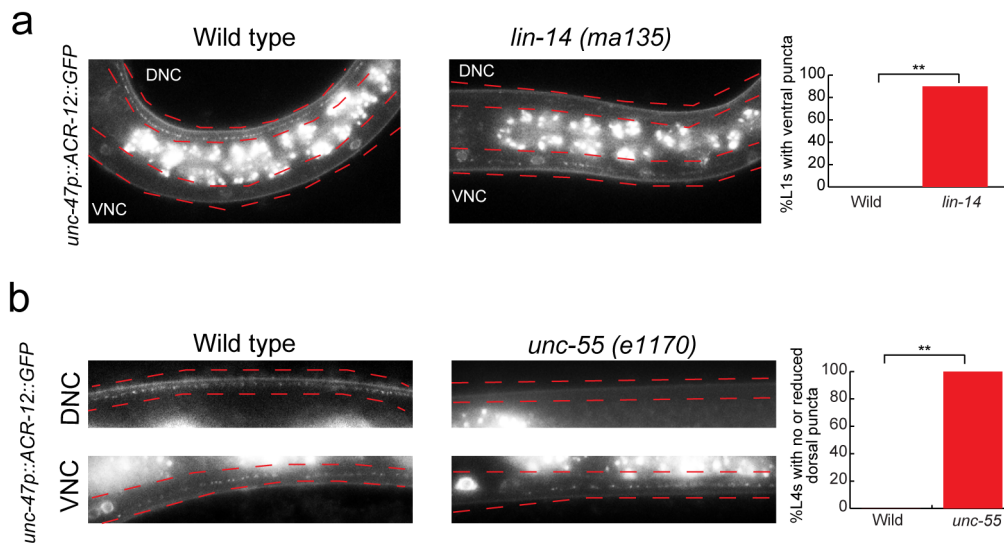
**Extended Data Figure 4 | Deletion of a putative UNC-30 binding site results in loss of *oig-1* expression in the D-type neurons.** Regions of the *oig-1* promoter were fused to *gfp* to analyse expression. (+) indicates robust expression of the reporter construct in the specified cell type, whereas (-) indicates loss of expression in the specified cell type. Twenty worms at both the L1 and L4 stage were scored for each line. Expression of a 1 kb promoter reporter (prom 1) recapitulates expression of the *oig-1<sup>fosmid</sup>::gfp* reporter in the D-type MNs (see Fig. 2). This region contained 3 elements that exactly match the UNC-30 consensus binding site (TAATC, purple box<sup>14</sup>) and multiple others that are a partial match to the UNC-30 binding site (magenta and blue boxes). Further deletion of this prom 1 defined a minimal 125 bp element that is sufficient to drive *oig-1* expression in the D-type MNs (prom 6). This

element contains two sites that partially match the UNC-30 binding sequence. Deletion of the AAATC site in the context of the 1 kb promoter (prom 9) has no effect on *oig-1* expression in the D-type MNs. Deletion of the TAAAC site in the context of the 1 kb promoter reporter (prom 10) results in complete loss of *oig-1* expression specifically in the D-type MNs. We noted that some of the smaller transcriptional reporter lines show extended expression of *gfp* in the DD motor neurons. Since DD rewiring is delayed upon partial removal of the homeobox gene *irx-1* (ref. 13), it is possible that in these reporters, potential IRX-1 binding sites are deleted. We have not pursued the effect of *irx-1* on *oig-1* expression as the lethality associated with complete loss of *irx-1* function complicates an analysis of *irx-1* null mutant phenotypes in D-type motor neurons.



**Extended Data Figure 5 | *oig-1* mutants defects and their rescue.** **a**, *oig-1* mutants display locomotory defects. Locomotion of L4 animals was analysed with tracking assays<sup>28</sup>. The graphs on the left side of each panel correspond to assays comparing wild-type, *oig-1* mutant, and *oig-1; unc-30p::oig-1* animals. The graphs on the right side of each panel correspond to assays comparing wild-type, *oig-1* mutant, and *oig-1; oig-1<sup>fosmid</sup>::gfp* animals. Twenty animals (each dot on a plot) were tracked for each genotype for both comparisons. Mean and Q values are indicated. Note that in a previously published analysis of a large panel of available mutants, the same set of locomotory defects that we describe here for *oig-1* mutants were found to be affected in *unc-55* mutants<sup>28</sup>, albeit in a stronger manner than *oig-1* mutants. Also note that the very strong locomotory defects *unc-30* defects are qualitatively very different from *oig-1* mutants, but this is to be expected as *unc-30* mutants do not only show the synaptic defects that we describe here, but also lack the neurotransmitter GABA (ref. 3), thereby disabling any neuromuscular signalling. Top, The midbody speed of *oig-1* mutant animals is significantly lower than that of wild-type animals. This defect is partially rescued (statistically different from *oig-1* mutants but also from wild-type animals) by expressing *unc-30p::oig-1* in *oig-1* animals (left graph). The lower midbody speed of *oig-1* mutants is completely rescued (statistically different from *oig-1* mutants but not from wild-type animals) by expressing the *oig-1<sup>fosmid</sup>::gfp* in *oig-1* mutants. Middle, *oig-1* mutants exhibit more dwelling than wild-type L4 animals. This defect is partially rescued (statistically different from *oig-1* mutants but also from wild-type animals) by expressing *unc-30p::oig-1* in *oig-1* animals (left graph).

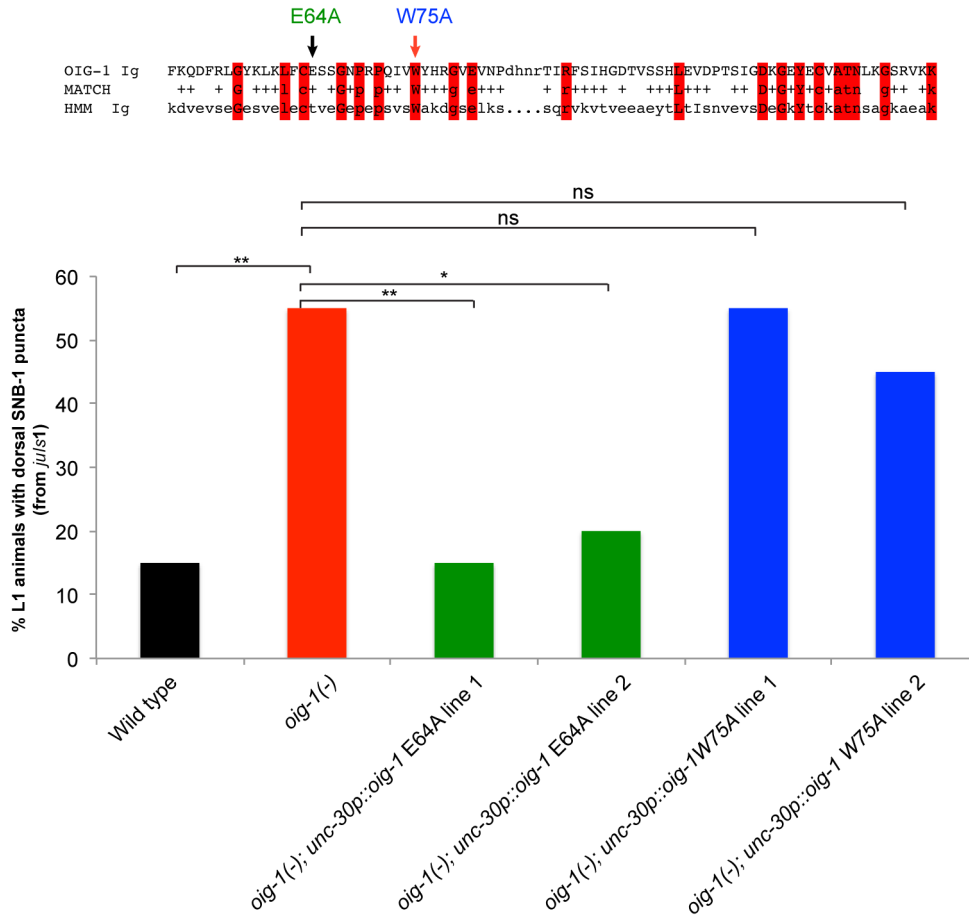
The increased dwelling of *oig-1* mutants is completely rescued (statistically different from *oig-1* mutants but from not wild-type animals) by expressing the *oig-1<sup>fosmid</sup>::gfp* in *oig-1* mutants. Bottom, *oig-1* mutants exhibit an increased path curvature compared to wild-type animals. This defect is not rescued by expressing *unc-30p::oig-1* in *oig-1* animals (left graph). The increased path curvature of *oig-1* mutants is completely rescued (statistically different from *oig-1* mutants but not from wild-type animals) by expressing the *oig-1<sup>fosmid</sup>::gfp* in *oig-1* mutants. **b**, Aldicarb-sensitivity defects in *oig-1* mutants. *oig-1* mutant young adult animals (red squares), which display aberrant GABAergic synapses in both the ventral and dorsal cord, show hypersensitivity to aldicarb-induced paralysis compared to wild type (black triangles). Expression of *oig-1<sup>fosmid</sup>::gfp* from a multicopy transgenic array (green circles) does not only rescue the *oig-1* mutant phenotype, but even results in a slight hyposensitivity to aldicarb. Worms (blinded for genotype) were tested every 15 min for paralysis by touching the head and tail three times each.  $n = 20$  for each strain, repeated 3 times. **c**, Expression of *oig-1* in the D-type neurons rescues ectopic DD synapses in the dorsal nerve cord. At the L1 stage when only the DD MNs are present, SNB-1-GFP (from *juIs1-unc-25p::snb-1::gfp*) localizes to the ventral nerve cord (VNC) in wild-type animals (top left). Ectopic SNB-1-GFP puncta localize to the dorsal nerve cord (DNC) of *oig-1* mutant L1s (top right). This phenotype is rescued by expressing *oig-1* in the D-type MNs (*unc-30p::oig-1*, *otEx4955*, bottom left), but not by expressing *oig-1* in the neighbouring cholinergic MNs (*unc-3p::oig-1*, *otEx4942*, bottom right). Original magnification,  $\times 630$ ; white boxes indicate the dorsal nerve cord.



**Extended Data Figure 6 | ACR-12 is mislocalized in *lin-14* and *unc-55* mutants.** Cholinergic innervation to the D-type MNs is visualized with an *unc-47p::acr-12::gfp* reporter transgene, maintained in an *acr-12(ok367)* mutant background<sup>17</sup>. **a**, ACR-12 puncta localization is affected by loss of *lin-14*. In wild-type L1 animals, ACR-12 puncta are observed only in the DD neurons in the dorsal nerve cord (DNC) (left). In *lin-14* mutant L1 animals (scored in progeny from *lin-14* null animals carrying a *lin-14* rescue array<sup>23</sup>), ACR-12 puncta are detected in the ventral nerve cord (VNC) of the DD MNs. Quantification of this data are represented in the graph. Some dorsal puncta in the DD MNs were still observed in 83% of the *lin-14* mutant L1s that had

puncta in the ventral nerve cord. L1 animals were obtained by hypochlorite-treating gravid adult animals and letting embryos hatch and arrest in M9 for 16–18 h.  $n > 20$  for each strain scored,  $**P < 0.01$ . **b**, ACR-12 puncta localization is affected by loss of *unc-55*. In wild-type L4 animals, ACR-12 puncta are observed in both the ventral (VNC) and dorsal (DNC) nerve cords (left). In *unc-55* mutant L4 animals, ACR-12 puncta are observed mostly in the ventral nerve cord of *unc-55* mutants. Ventral and dorsal nerve cords are marked by red dotted lines. Quantification of this data are represented in the graph.  $n > 20$  for each strain,  $**P < 0.01$ , original magnification,  $\times 630$ .

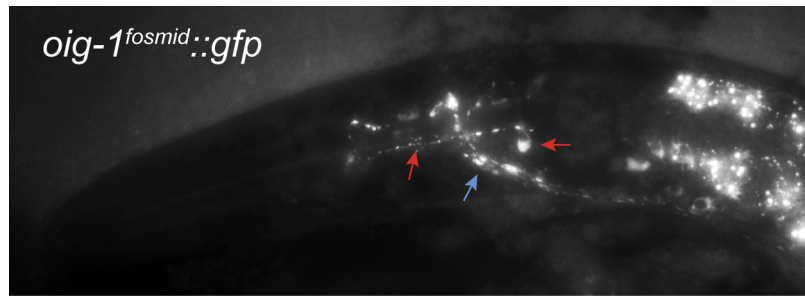




**Extended Data Figure 7 | The IgC2 domain is necessary for OIG-1 function.**

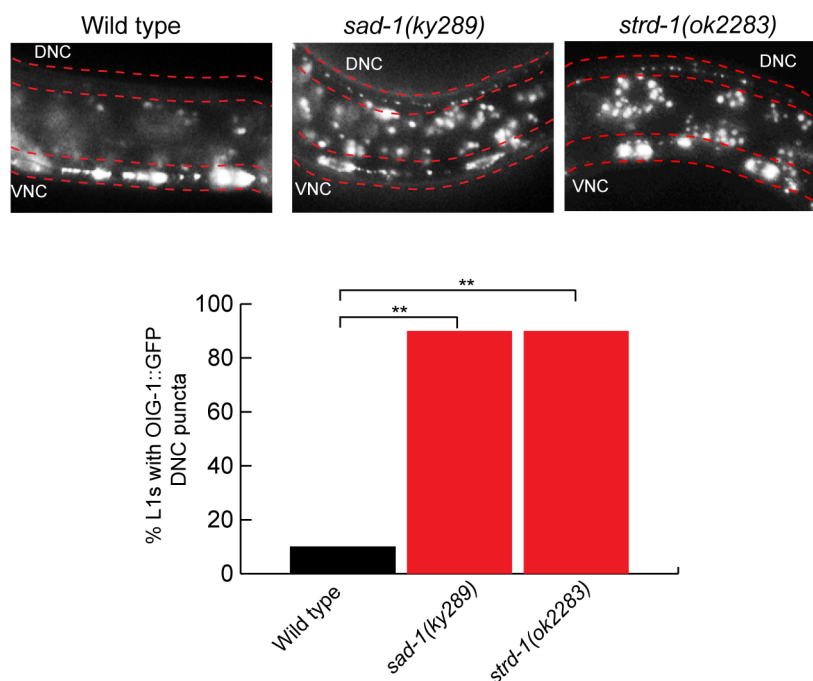
At the L1 stage when only the DD MNs are present, ectopic SNB-1-GFP puncta (from *juls1-unc-25p::SNB-1::GFP*) localize to the dorsal nerve cord (DNC) of *oig-1* mutant L1s (red bar) but not in wild-type animal (black bar) (see Fig. 3a). Based on an alignment with the hidden Markov model (HMM) Ig domain (top), a highly conserved residue (W75) and a nonconserved residue (E64) in the OIG-1 Ig domain were mutated in the context of an *unc-30p::oig-1*

transgene that is able to rescue the L1 ectopic synapse defects (see Fig. 3a). The *unc-30p::oig-1E64A* transgenes (*otEx6212*, *otEx6213*) were still able to rescue the synaptic defects of *oig-1* L1 animals (green bars), whereas the *unc-30p::oig-1W75A* transgenes (*otEx6214*, *otEx6215*) had no rescue ability (blue bars). L1 animals were obtained by hypochlorite-treating gravid adult animals and letting embryos hatch and arrest in M9 for 16–18 h.  $n > 20$  for each strain scored,  $**P < 0.01$ ,  $*P < 0.05$ .



**Extended Data Figure 8 | OIG-1 localization in other neuron types.** OIG-1–GFP (from *oig-1<sup>fosmid</sup>::gfp*) localizes in a punctate manner along axons in the nerve ring (blue arrow) and along a pair of neurons in the pharynx, tentatively identified as the M2 MNs (red arrows point to cell body and process). These

neurons form synapses onto pharyngeal muscles along their processes, and these processes also show punctate localization of OIG-1. Original magnification,  $\times 630$ .



**Extended Data Figure 9 | OIG-1 is mislocalized in *sad-1* and *strd-1* mutants.** OIG-1-GFP (from *oig-1<sup>fosmid</sup>::gfp*) localizes to the ventral nerve cord (VNC) of wild-type L1 animals (left). In *sad-1* mutants (middle), OIG-1-GFP is ectopically localized to the dorsal side (DNC) of L1 animals. In *strd-1* mutants,

OIG-1-GFP is ectopically localized to the dorsal side of L1 animals. Quantification of the data are shown in graph.  $n > 20$  for each strain scored,  $**P < 0.01$ , original magnification,  $\times 630$ .



# Cell-intrinsic adaptation of lipid composition to local crowding drives social behaviour

Mathieu Frechin<sup>1</sup>, Thomas Stoeger<sup>1,2</sup>, Stephan Daetwyler<sup>1</sup>, Charlotte Gehin<sup>3</sup>, Nico Battich<sup>1,2</sup>, Eva-Maria Damm<sup>4</sup>, Lilli Stergiou<sup>4</sup>, Howard Riezman<sup>3</sup> & Lucas Pelkmans<sup>1</sup>

Cells sense the context in which they grow to adapt their phenotype and allow multicellular patterning by mechanisms of autocrine and paracrine signalling<sup>1,2</sup>. However, patterns also form in cell populations exposed to the same signalling molecules and substratum, which often correlate with specific features of the population context of single cells, such as local cell crowding<sup>3</sup>. Here we reveal a cell-intrinsic molecular mechanism that allows multicellular patterning without requiring specific communication between cells. It acts by sensing the local crowding of a single cell through its ability to spread and activate focal adhesion kinase (FAK, also known as PTK2), resulting in adaptation of genes controlling membrane homeostasis. In cells experiencing low crowding, FAK suppresses transcription of the ABC transporter A1 (ABCA1) by inhibiting FOXO3 and TAL1. Agent-based computational modelling and experimental confirmation identified membrane-based signalling and feedback control as crucial for the emergence of population patterns of ABCA1 expression, which adapts membrane lipid composition to cell crowding and affects multiple signalling activities, including the suppression of ABCA1 expression itself. The simple design of this cell-intrinsic system and its broad impact on the signalling state of mammalian single cells suggests a fundamental role for a tunable membrane lipid composition in collective cell behaviour.

Adherent tissue culture cells spread out their cell surface more when experiencing low local crowding than high local crowding, resulting in a higher number of focal adhesions, sites of cellular attachment to the extracellular matrix (ECM), and higher levels of activated FAK (Extended Data Fig. 1a). FAK is recruited to focal adhesions, where it undergoes autophosphorylation, and subsequently recruits and phosphorylates phosphatidylinositol-3-OH kinase (PI(3)K) and many other proteins involved in signalling, cell adhesion and cytoskeletal dynamics<sup>4–6</sup>. FAK may thus, in a cell-intrinsic manner, sense local cell crowding by reacting to the available space and mechanical constraints imposed during cell population growth<sup>7,8</sup>, and signal this to downstream cellular functions. To test this, we compared the extent of adaptation of the transcriptome to cellular crowding in adherent embryonic fibroblasts from a FAK-knockout mouse (FAK-KO) with cells from the same background in which FAK was stably re-expressed (FAK-rescue).

A total of 1,014 genes (~5% of the whole genome) adapt their transcript abundance to cellular crowding, of which 80% required the presence of FAK to adapt (Fig. 1a). Although FAK induces genes related to cell growth and proliferation (Extended Data Fig. 1b), it suppresses genes involved in membrane and organelle homeostasis (Fig. 1b) in cells experiencing low crowding, amongst which are 4 ATP-binding cassette (ABC) transporters (*Abca1*, *Abca6*, *Abca9* and *Abcg2*) (Extended Data Fig. 1c). *Abca1* was the overall second most strongly suppressed (~14-fold) gene by FAK (Fig. 1a) and the strongest hit amongst all genes in functional annotation terms related to membrane organization (Fig. 1b). ABC transporters mediate

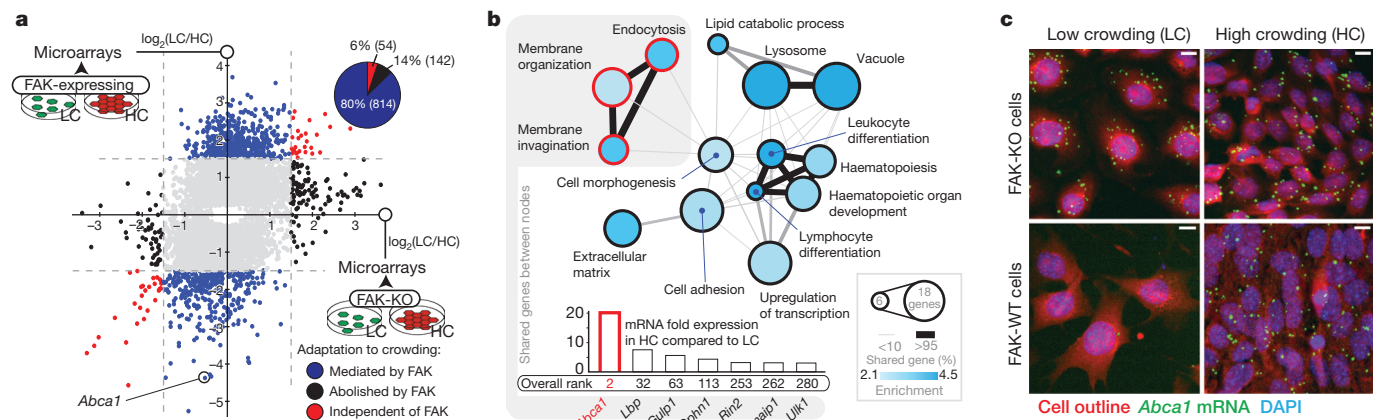
the transport of various substrates across membranes, including phospholipids and cholesterol<sup>9,10</sup>.

Single-molecule fluorescence *in situ* hybridization and automated image analysis<sup>3,11</sup> confirmed the transcriptomics results at the single-cell level, showing that FAK controls the abundance of *Abca1* transcripts in single cells to local crowding (Fig. 1c and Extended Data Fig. 1d, e). This adaptation involves low (1–20) and highly variable transcript copy numbers (Extended Data Fig. 1d), and also occurs in the presence of growth factors and cytokines in the medium (Extended Data Fig. 1f).

Predicted candidate transcription factors (see Supplementary Information and Supplementary Table 2) were tested for their involvement in this adaptation using RNA-mediated interference (RNAi) in cells that lack FAK (FAK-KO) and thus highly express *Abca1* independent of crowding. RNAi of *Foxo3*, *Tal1* and *Stat4*, as well as *Lxrb* (liver X receptor beta, also known as *Nr1h2*), the canonical transcription factor driving expression of ABCA1 (ref. 12), reduced *Abca1* transcript abundance in these cells by ~50% (Extended Data Fig. 2a). As TAL1 and FOXO3 are phosphorylated by the serine/threonine kinase AKT, which is activated by PI(3)K downstream of FAK<sup>5</sup>, leading to rapid degradation of TAL1 (ref. 13) and inactivation of FOXO3 (ref. 14), we focused on these transcription factors. Chromatin immunoprecipitation (ChIP) experiments (Extended Data Fig. 2b) revealed that in cells lacking FAK, both FOXO3 and TAL1 bind to *Abca1* chromatin independent of cellular crowding. In cells expressing FAK, FOXO3 and TAL1 bind to *Abca1* chromatin at closely located positions only when cells experience high crowding (Fig. 2a). This is in contrast to LXRβ, which constitutively binds to *Abca1* chromatin independent of cellular crowding or the presence of FAK (Fig. 2a). Furthermore, western blots of multiple adherent cell lines revealed that cells experiencing low crowding contain higher levels of phosphorylated PI(3)K, AKT and FOXO3 and lower levels of TAL1 than cells experiencing high crowding. Consequently, these cells express a low amount of ABCA1 protein at low cellular crowding. Inhibition of PI(3)K (by wortmannin or LY-294002) lack of FAK (FAK-KO), or inhibition of FAK (by Y15) abolished these differences, leading to ABCA1 expression also in cells experiencing low crowding (Fig. 2b–e and Extended Data Fig. 2c–e). These effects were observed in mouse embryonic fibroblasts, human lung epithelial cells and freshly isolated human keratinocytes. Micropatterns confirmed that cell crowding-dependent expression of ABCA1 stems from the available space of a single cell to adhere to, consistent with a cell-intrinsic mechanism of adaptation (Extended Data Fig. 2f).

To understand if this cell-intrinsic mechanism can drive multicellular pattern formation, we applied single-cell mathematical modelling and computer simulation using a coupled two-level agent-based modelling<sup>15</sup> and differential equation approach (Supplementary Information (mathematical appendix)). The agent-based model simulates the dynamic behaviour of focal adhesions (Supplementary Video 1) and their adhesion potential in multiple single cells of a

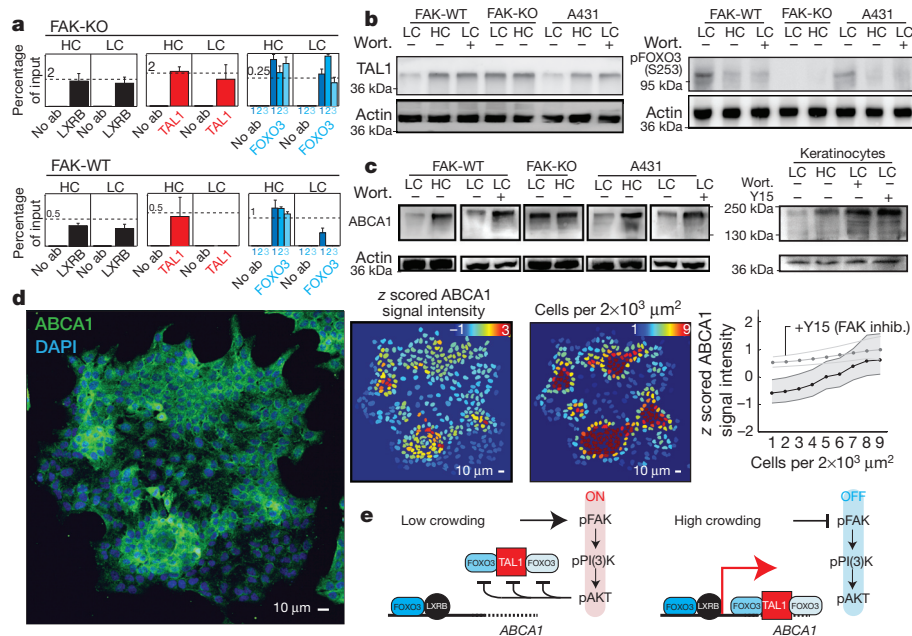
<sup>1</sup>Faculty of Sciences, Institute of Molecular Life Sciences, University of Zurich, 8057 Zurich, Switzerland. <sup>2</sup>Life Science Zurich Graduate School, Ph.D. program in Systems Biology, ETH Zurich and University of Zurich, 8057 Zurich, Switzerland. <sup>3</sup>Department of Biochemistry, University of Geneva, 1205 Geneva, Switzerland. <sup>4</sup>Institute of Molecular Systems Biology, ETH Zurich, 8057, Zurich, Switzerland.



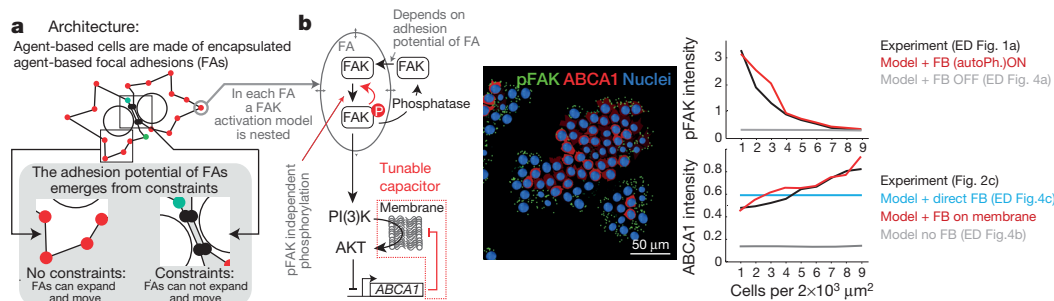
crowding. Node colour: enrichment, node size: number of genes, edge width: number of overlapping genes between nodes. c, Branched DNA (bDNA) single-molecule FISH against *Abca1* transcripts in FAK-KO (representative of  $1.2 \times 10^4$  cells) or FAK-WT (representative of  $1.5 \times 10^4$  cells) MEFs experiencing low or high crowding. DAPI, 4',6-diamidino-2-phenylindole. Scale bars, 10  $\mu$ m.

Supplementary Video 3). Timescale coupling could be achieved by the property of the membrane to act as a storage for phosphatidylinositol-3,4,5-triphosphate (PtdIns(3,4,5)P<sub>3</sub> or PIP<sub>3</sub>) production by PI(3)K, while adaptation may be achieved by the ability of ABCA1 to alter physical properties of the membrane<sup>16</sup> leading to a decreased lipid ordering and increased diffusion rate of lipids<sup>17</sup>, which affects the probability of AKT activation on the membrane by phosphoinositide-dependent kinase 1 (PDK1)<sup>18,19</sup>. We thus modelled the membrane as a ‘tunable capacitor’ (Fig. 3b and Extended Data Fig. 3e) that stores PIP<sub>3</sub> and that can be perturbed by ABCA1 in its capacity to activate AKT. This generates a pattern of ABCA1 expression similar to experimental observations that is insensitive to fluctuations in most parameters and primarily depends on the strength of ABCA1 feedback (Extended Data Fig. 6b–d). It also recapitulates the dynamics of ABCA1 down-regulation in scratch assays, when cells at high local crowding suddenly become exposed to free space (Extended Data Fig. 4f, g).

To investigate the existence of ABCA1 feedback on the capacitor function of the membrane, we examined whether the naturally observed crowding-dependent cell-to-cell variability in ABCA1



**a**, ChIP of *Abca1* ( $n = 3$  biological replicates, each mean of 3 technical replicates, s.d.) in FAK-KO and FAK-WT MEFs at low or high crowding. No ab, no antibody. **b**, Western blots of pFOXO3 and TAL1 levels in FAK-WT and FAK-KO MEFs, and A431 (human epidermoid carcinoma) cells at low crowding, high crowding, or low crowding + wortmannin. **c**, Western blots of ABCA1 levels as above, including in primary human keratinocytes and upon treatment with Y15 (FAK inhibitor). **d**, Immunofluorescence imaging of ABCA1 in a population of A431 cells (left, representative of  $10^4$  cells), quantified single-cell intensities of ABCA1 staining (centre) and local cell crowding (right). Curves, single-cell ABCA1 intensities against local cell crowding with or without Y15. Interquartile area in grey,  $n$  (single cells)  $> 10^4$ . **e**, Diagram of the FAK-ABCA1 pathway at low and high crowding. pFAK, phosphorylated FAK.



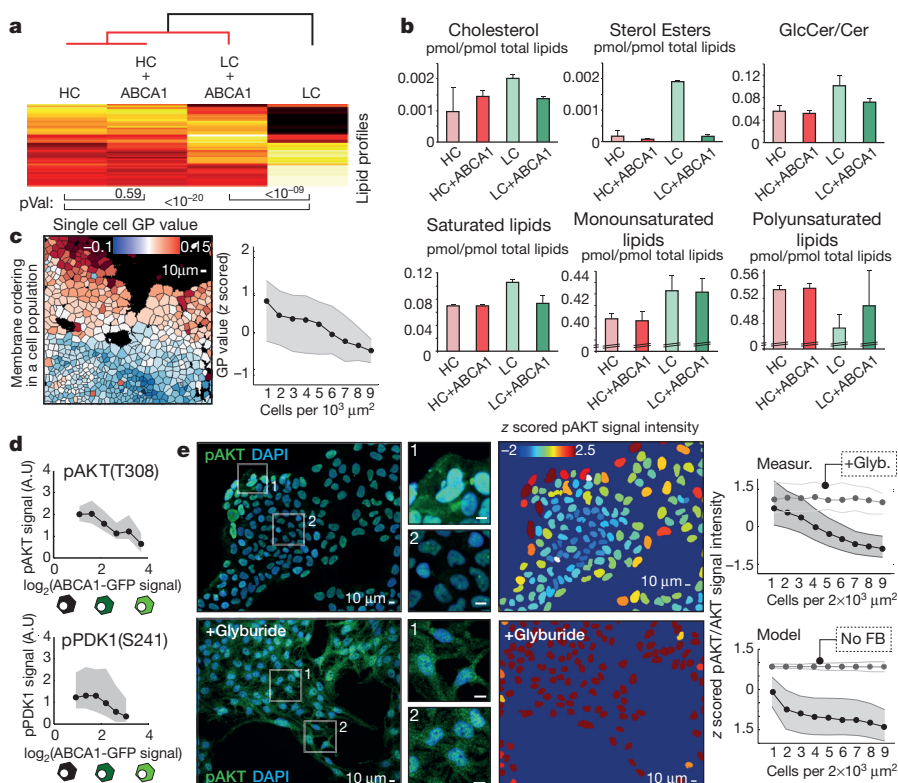
**Figure 3 | Multi-scale model of the FAK-ABCA1 system.** **a**, Architecture of agent-based modelled single cells encapsulating multiple agent-based modelled focal adhesions. **b**, Model of FAK activation nested in each focal adhesion, influenced by the adhesion potential of each focal adhesion emerging from **a** (left, top part). Model-simulated pFAK levels in single cells (centre image, green signal, representative of all simulations using the same parameters, this run:  $10^3$  cells) and quantification (right, top graph) against local cell crowding without (grey, Extended Data Fig. 4a) and with (red) positive feedback (FB),

expression causes changes in membrane lipid composition. Cells experiencing high crowding have a strikingly different lipid composition than cells experiencing low crowding (Fig. 4a and Supplementary Table 3). Cells experiencing low crowding which expressed ABCA1 at levels naturally found in cells experiencing high crowding. (Extended Data Fig. 7a) have a lipid composition more closely resembling that of cells experiencing high crowding (Fig. 4a and Extended Data Fig. 7b). In particular, cells at low crowding have a higher amount of free cholesterol, higher levels of cholesteryl esters (Fig. 4b and Extended Data Fig. 7c), more lipid droplets (Extended Data Fig. 7f), a higher ratio of glucosylceramide over ceramide (GlcCer/Cer) (indicative of glycosphingolipid biosynthesis rate), higher levels of saturated lipids, and lower levels of monounsaturated and polyunsaturated lipids than cells at high crowding (Fig. 4b and Extended Data Fig. 7d, e). In cells experiencing high crowding, plasmid-driven expression of ABCA1 did

experiments in black (Extended Data Fig. 1a). Control of ABCA1 transcription by FAK using a tunable membrane capacitor topology, which involves PI(3)K and AKT and feedback by ABCA1 (left, bottom part). Model-simulated ABCA1 levels in single cells (centre, red signal), and quantification (right, bottom graph) against local cell crowding without feedback (grey, Extended Data Fig. 4b), with direct feedback (light blue, Extended Data Fig. 4c), and with tunable capacitor (red). Experiments in black.

not alter lipid composition (Fig. 4a, b, and Extended Data Fig. 7b–e). As a consequence, cells experiencing high crowding display lower membrane lipid ordering than cells experiencing low crowding (Fig. 4c), mediated by the crowding-dependent expression of ABCA1 (Extended Data Fig. 7g). Cells that lack FAK and thus express high levels of ABCA1 contain less cholesterol and less of the glycosphingolipid GM1 and display lower membrane lipid ordering than cells expressing FAK (Extended Data Fig. 7h, i).

Similarly, we found that ABCA1 levels influence the amount of S241-phosphorylated PDK1 and T308-phosphorylated AKT (Fig. 4d). Accordingly, levels of T308-phosphorylated AKT are higher in cells experiencing low crowding than cells experiencing high crowding (Fig. 4e). Pharmacological inhibition of ABCA1 abolished this pattern, increasing the level of T308-phosphorylated AKT in cells experiencing high crowding, as predicted by the model when the



**Figure 4 | The FAK-ABCA1 system adapts membrane lipid composition, ordering and signalling to local crowding.** **a**, Hierarchical clustering of lipid profiles, see Extended Data Fig. 6b and Supplementary Table 3. **b**, Histograms of selected lipid species (for free cholesterol in nmol per cell, see Extended Data Fig. 7c). For *P* values (*t*-test), see Extended Data Fig. 7d ( $n = 4$  biological replicates, each the mean of 4 technical replicates, s.d.). **c**, Z-scored general polarization (GP) values (see Extended Data Fig. 7g) per single A431 cells (left) stained with Laurdan against local cell crowding (right) (interquartile area in grey, number of single cells  $> 3 \times 10^3$ ). **d**, The effect of levels of ABCA1-GFP, randomly expressed from a plasmid in A431 cells at low crowding on pAKT and pPDK1 in single cells (interquartile area in grey). **e**, Untreated (top panels) or glyburide-treated (bottom panels) A431 cells immunostained against pAKT (T308). Nucleus segmentation images are colour-coded for pAKT levels. Top curves (left): single-cell pAKT levels against local crowding in absence (grey) or presence of glyburide (white) ( $n$  single cells  $> 10^4$ ). Bottom curves: model-predicted pAKT levels against local crowding with (grey) or without (white) feedback (interquartile areas in grey).



double-negative feedback is removed (Fig. 4e). In addition, exogenous loading of the membrane with cholesterol and the glycosphingolipid GM1, as well as pharmacological inhibition of ABCA1, increases the level of phosphorylated PDK1 and AKT in cells lacking FAK (Extended Data Fig. 7j). Thus, ABCA1 inhibits the FAK-induced signalling pathway that suppresses its own transcription by adapting membrane lipid composition, confirming the membrane-based feedback predicted by the model as a requirement for gradual patterning. We made similar observations for levels of phosphorylated STAT3 and PAK1/2, which are respectively an effector of cytokine receptors and of the small GTPase RAC1, both sensitive to membrane lipid composition (Extended Data Fig. 8)<sup>20,21</sup>. This indicates that the adaptation of membrane lipid composition to local crowding by the FAK–ABCA1 system influences multiple signalling pathways in cells, including those involved in cell motility and paracrine signalling.

We have uncovered a cell-intrinsic molecular mechanism that allows patterning of membrane lipid composition and signalling according to local crowding in a cell population. Several genes with roles in membrane homeostasis may participate in this patterning system, including multiple ABC transporters and lipid-processing enzymes (see Supplementary Table 1, Extended Data Fig. 9 and Supplementary Discussion). In our minimal model, pattern formation of membrane lipid composition only requires variation in the extent of cellular crowding to emerge as cells proliferate. Patterning is subsequently promoted and stabilized by feedback loops without the need for specific cell–cell communication. Because lipid composition affects many membrane protein activities, adapting it to local crowding may have a fundamental role in controlling cellular behaviour within a social context, from colony formation in unicellular organisms<sup>22</sup> to collective cell migration<sup>23</sup>, haematopoiesis<sup>24</sup> and T cell activation<sup>25</sup>, and the control of epithelial cell proliferation in multicellular organisms<sup>26</sup>.

Our work indicates a crucial role for membrane-based signalling in this cell-intrinsic system, in which the membrane may act as a capacitor that converts signals to the correct timescale and is tuned by enzymes that alter membrane lipid composition and ordering in a feedback mechanism. Both timescale adaptation and feedback are required for gradual patterns in a growing cell population to emerge. It will now be important to unravel how such a tunable capacitor operates mechanistically, and to generalize this concept to the possible uses of cellular structures in signal computation.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 25 September 2014; accepted 25 March 2015.**

**Published online 25 May 2015.**

- Kicheva, A., Cohen, M. & Briscoe, J. Developmental pattern formation: insights from physics and biology. *Science* **338**, 210–212 (2012).
- Tabata, T. Genetics of morphogen gradients. *Nature Rev. Genet.* **2**, 620–630 (2001).
- Snijder, B. *et al.* Population context determines cell-to-cell variability in endocytosis and virus infection. *Nature* **461**, 520–523 (2009).
- Guan, J. L. & Shalloway, D. Regulation of focal adhesion-associated protein tyrosine kinase by both cellular adhesion and oncogenic transformation. *Nature* **358**, 690–692 (1992).
- Schaller, M. D. Cellular functions of FAK kinases: insight into molecular mechanisms and novel functions. *J. Cell Sci.* **123**, 1007–1013 (2010).

- Mitra, S. K., Hanson, D. A. & Schlaepfer, D. D. Focal adhesion kinase: in command and control of cell motility. *Nature Rev. Mol. Cell Biol.* **6**, 56–68 (2005).
- Puliafito, A. *et al.* Collective and single cell behavior in epithelial contact inhibition. *Proc. Natl Acad. Sci. USA* **109**, 739–744 (2012).
- Piccolo, S., Dupont, S. & Cordenonsi, M. The biology of YAP/TAZ: Hippo signaling and beyond. *Physiol. Rev.* **94**, 1287–1312 (2014).
- Tarling, E. J., Vallim, T. Q. D. A., Edwards, P. & a. Role of ABC transporters in lipid transport and human disease. *Trends Endocrinol. Metab.* **24**, 342–350 (2013).
- Lawn, R., Wade, D. & Garvin, M. The Tangier disease gene product ABC1 controls the cellular apolipoprotein-mediated lipid removal pathway. *J. Clin. Invest.* **104**, 25–31 (1999).
- Battich, N., Stoeger, T. & Pelkmans, L. Image-based transcriptomics in thousands of single human cells at single-molecule resolution. *Nature Methods* **10**, 1127–1133 (2013).
- Costet, P., Luo, Y., Wang, N. & Tall, A. R. Sterol-dependent transactivation of the ABC1 promoter by the liver X receptor/retinoid X receptor. *J. Biol. Chem.* **275**, 28240–28245 (2000).
- Palamarchuk, A. *et al.* Akt phosphorylates Tal1 oncoprotein and inhibits its repressor activity. *Cancer Res.* **65**, 4515–4519 (2005).
- Brunet, A. *et al.* Akt promotes cell survival by phosphorylating and inhibiting a forkhead transcription factor. *Cell* **96**, 857–868 (1999).
- Holcombe, M. *et al.* Modelling complex biological systems using an agent-based approach. *Integr. Biol. (Camb)* **4**, 53–64 (2012).
- Zarubica, A. *et al.* Functional implications of the influence of ABCA1 on lipid microenvironment at the plasma membrane: a biophysical study. *FASEB J.* **23**, 1775–1785 (2009).
- Saffman, P. G. & Delbrück, M. Brownian motion in biological membranes. *Proc. Natl Acad. Sci. USA* **72**, 3111–3113 (1975).
- Lasserre, R. *et al.* Raft nanodomains contribute to Akt/PKB plasma membrane recruitment and activation. *Nature Chem. Biol.* **4**, 538–547 (2008).
- Landry, Y. D. *et al.* ATP-binding cassette transporter A1 expression disrupts raft membrane microdomains through its ATPase-related functions. *J. Biol. Chem.* **281**, 36091–36101 (2006).
- Shah, M., Patel, K., Fried, V. A. & Sehgal, P. B. Interactions of STAT3 with caveolin-1 and heat shock protein 90 in plasma membrane raft and cytosolic complexes: preservation of cytokine signaling during fever. *J. Biol. Chem.* **277**, 45662–45669 (2002).
- del Pozo, M. A. *et al.* Integrins regulate Rac targeting by internalization of membrane domains. *Science* **303**, 839–842 (2004).
- Vlamakis, H., Chai, Y., Beauregard, P., Losick, R. & Kolter, R. Sticking together: building a biofilm the *Bacillus subtilis* way. *Nature Rev. Microbiol.* **11**, 157–168 (2013).
- Friedl, P. & Gilmour, D. Collective cell migration in morphogenesis, regeneration and cancer. *Nature Rev. Mol. Cell Biol.* **10**, 445–457 (2009).
- Yvan-Charvet, L. *et al.* ATP-binding cassette transporters and HDL suppress hematopoietic stem cell proliferation. *Science* **328**, 1689–1693 (2010).
- Bensinger, S. J. *et al.* LXR signaling couples sterol metabolism to proliferation in the acquired immune response. *Cell* **134**, 97–111 (2008).
- Lee, B. H. *et al.* Dysregulation of cholesterol homeostasis in human prostate cancer through loss of ABCA1. *Cancer Res.* **73**, 1211–1218 (2013).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank B. Snijder for help with single-cell Laurdan quantification, P. Liberali for help with imaging of cell-to-cell variability, Y. Yakimovich for IT infrastructure support, and all members of the laboratory for discussions and support. M.F. was supported by an EMBO and a Marie Curie (301650) fellowship. E.-M.D. was supported by an Oncosuisse fellowship. L.S. was supported by a Bonizzi Theler fellowship. This work is supported by the University of Zurich and the SystemsX.ch RTD Project LipidX.

**Author Contributions** L.P. supervised and conceived the project, M.F., T.S., E.-M.D. and L.S. performed experiments, C.G. and H.R. performed lipid mass spectrometry, M.F. and N.B. developed computational image analysis methods, M.F. and L.P. performed data analysis, M.F. and S.D. developed mathematical models, M.F. performed mathematical modelling, L.P. and M.F. wrote the manuscript.

**Author Information** The microarray data set has been uploaded to the NCBI Gene Expression Omnibus as record GSE43873. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L.P. ([lucas.pelkmans@imls.uzh.ch](mailto:lucas.pelkmans@imls.uzh.ch)).

## METHODS

**Cell culture.** Media and reagents were from GibcoBRL. Wild-type MEFs (FAK-WT), or knockout for FAK (FAK-KO), and A431 cells were purchased from ATCC. Mouse embryonic fibroblasts rescued for FAK (FAK-rescue) were a gift from C. Hauck (University of Konstanz, Germany). E. Reichmann and L. Pontiggia provided keratinocyte primary cells (UZH, Zurich). Standard growth conditions were the following, cells were incubated 3 to 4 days using DMEM containing 10% FBS and  $1 \times$  glutamine ( $+135 \mu\text{g ml}^{-1}$  hygromycinB for the FAK-rescue cells) at  $37^\circ\text{C}$  under 5%  $\text{CO}_2$ . Initial cell number was  $2 \times 10^5$  to  $2.5 \times 10^5$  cells for 10-cm dishes  $3 \times 10^4$  to  $5 \times 10^4$  cells per well for 12 wells plates containing 13 mm coverslips and  $2 \times 10^3$  to  $2.5 \times 10^3$  cells per well for 96-well plates. All our cell lines are tested on a monthly basis for mycoplasma contamination using chemiluminescent assay. The service is independent, centralized for all the UZH and provided at the institute of virology of the UZH. Once the desired population pattern is reached (see video in ref. 3, Snijder *et al.* 2009) cells are serum deprived for approximately 12 h and used for subsequent preparations. Wortmannin (100 nM), Y15 (25  $\mu\text{M}$ ), LY-294002 (10  $\mu\text{M}$ ) and glyburide (25  $\mu\text{M}$ ) treatments were performed over approximately 12 h before preparation. Coverslips were mounted on glass slide using Immu-Mount (Thermo Scientific), a water-based mounting medium.

**Plasmid transfection.** FAK-WT cells grown in 96-well plates or 10-cm dishes were transfected respectively with 80 ng per well or 4  $\mu\text{g}$  of ABCA1 construct carried in the pEGFP-N1 backbone mixed with 0.2 or 10  $\mu\text{l}$  lipofectamine2000 following the manufacturer's specifications. Homo sapiens ABCA1 coding sequence was synthesized *de novo* and inserted between SacI and SacII restriction sites. The cloned ABCA1 sequence corresponds to the full-length consensus coding sequence CCDS6762.1.

**Cholesterol and GM1 staining.** Cells were quickly washed with successive  $1 \times$  PBS, 5% delipidated BSA,  $1 \times$  PBS and fixed for 4 min with 4% PFA. Cholesterol was stained using 0.01  $\text{mg ml}^{-1}$  filipin (Sigma) for 20 min, after two washes of 5 min in PBS, surface GM1 was stained using 0.2  $\mu\text{g ml}^{-1}$  cholera toxin subunit B (Alexa Fluor 555 conjugate, Invitrogen) for 10 min.

**Laurdan live staining.** Cells were grown in ibidi  $\mu$ -Slide 8 well chambers under standard conditions. Five minutes before acquisition, cells were mounted on the microscope (see microscope section) with environmental control and live stained by addition of 6-dodecanoyl-2-dimethylaminonaphthalene (Laurdan, Molecular Probes) and DraQ5 (Cell Signaling) at 5 and 0.5  $\mu\text{M}$  final concentrations directly in the medium. Images were acquired within the next 2 min.

**Immunostaining.** Unless specified, cells were grown following standard procedures. Fixation was performed with 4%PFA for 10 min, permeabilization with 0.1% Triton X-100 for 10 min, blocking with 1% BSA, 50 mM  $\text{NH}_4\text{Cl}$  for 30 min. Primary and secondary antibodies were diluted in blocking solution, treatments were separated by two 30-min PBS washes. Secondary antibody was applied for 1 h (Alexa Fluor 488 or 568 goat anti rabbit antibody, Invitrogen, 1  $\mu\text{g ml}^{-1}$ ). Nuclear staining is performed with 1  $\mu\text{M}$  DAPI for 10 min and cell outlines are visualized with Alexa Fluor 647 carboxylic acid succinimidyl ester (Life Science,  $10^{-4}$  dilution) staining for 10 min. For the pFAK staining, primary antibody was applied for 3 h (rabbit anti-pFAK (Y397) antibody, Cell Signaling no. 3283, 1:200) as well as for ABCA1 (rabbit anti-ABCA1 antibody, Abcam ab7360, 1:500). For pAKT (rabbit anti-pAKT (T308) antibody, Cell Signaling no. 2965, 1:1,000), pPDK1 (rabbit anti-pPDK1 (S241), no. 3061, Cell Signaling, 1:1,000), pSTAT3 (rabbit anti-pSTAT3 (T705) antibody, Cell Signaling no. 9131, 1:500) and pPAK1 (rabbit anti-pPAK1/2 (T423/T402) antibody, Cell Signaling no. 2601, 1:200) staining, primary antibody was applied overnight at  $4^\circ\text{C}$ .

**mRNA bDNA-FISH experiments.** FAK-WT cells were grown following standard conditions in 96-well plates. *Abca1* mRNA bDNA-FISH experiments and image based analysis were performed using the protocol and computational method published by our laboratory<sup>11</sup>. Briefly, cells were fixed, permeabilized, and protease K treated for the *Abca1* mRNA specific probe set to access properly its target sequences. A three-step treatment with successive pre-amplifier, amplifier and fluorescent probes hybridization allows the amplification of the mRNA probe signal and the visualization of single *Abca1* mRNAs. Nuclear staining was performed with 1  $\mu\text{M}$  DAPI for 10 min. Cell outlines were visualized with Alexa Fluor 647 carboxylic acid succinimidyl ester (Life Science) ( $10^{-4}$  dilution) staining for 10 min.

**Microscopes.** Laurdan, filipin and cholera toxin B images were acquired with  $40\times$  magnification on a Leica SP5 confocal microscope equipped with a UV laser ( $\lambda$ , 355 nm) in addition to the usual set of visible light lasers, for proper stimulation of Laurdan and filipin. Confocal images of pFAK were acquired on a Zeiss LSM710 microscope with  $40\times$  magnification (Zeiss NA1.2, C-apochromat, Korr UV-VIS-IR), GFP-FAK total internal reflection fluorescence (TIRF) video images were acquired on a Nikon visiView microscope with  $100\times$  magnification. Immunostainings of ABCA1, pS6, pAKT, pPI(3)K, pSTAT3, pPAK1 and mRNA

bDNA-FISH images were acquired on an automated Yokogawa CV7000 spinning disk microscope.

**Image analysis.** All image analysis was performed using CellProfiler<sup>27</sup> following the same procedure we used in previous publications<sup>3,11,28</sup>, with the help of additional MATLAB scripts published previously for the calculation of cellular crowding<sup>3</sup> or written specifically for this study for Laurdan image analysis (see specific section). The general image analysis pipeline was as follows. First, nuclei were detected and segmented based on the DAPI or DraQ5 stain using IdentifyPrimaryObjects CellProfiler module. Then, cell boundaries were estimated using nuclear propagation in IdentifySecondaryObjects CellProfiler module. Standard CellProfiler texture, intensity, size and shape features were extracted from nucleus and cell regions. We additionally implemented several image analysis steps for the purpose of detection of out of focus images and for the Support Vector Machine (SVM)-based classification<sup>29</sup> of poorly segmented nuclei.

**Membrane ordering analysis.** A dedicated CellProfiler module has been developed for this study (the code is available upon request) for defining automatically single-cell generalized polarization (scGP) values after nuclear and cell segmentation. This measurement is based on a previous publication<sup>30</sup> and works as follows: images of cells stained with Laurdan (see specific section above for details) are simultaneously acquired in the 400–460 nm (I1) and 470–530 nm (I2) wavelength windows after stimulation at 355 nm. The GP value is defined for each pixel following the formula:

$$\text{pxGP} = \frac{I1 - I2}{I1 + I2}$$

The mean GP value of each single cell (scGP value) is then defined by the mean of all pxGP values contained in each segmented cell.

**Microarray analysis.** High and low crowding FAK-rescue and FAK-KO cells were grown for 24 h in 10-cm dishes, in 10 ml of standard medium (described in the cell culture and preparation section). High crowding cells were seeded at a concentration of  $10^6$  cells per ml and low crowding cells at  $0.4 \times 10^5$  cells per ml. RNA preparations were done with the Qiagen RNeasy Mini Kit according to the manufacturer's manual, including the optional column DNase treatment.

The quality of the isolated RNA was determined with a NanoDrop ND 1000 (NanoDrop Technologies, Delaware, USA) and a Bioanalyzer 2100 (Agilent, Waldbronn, Germany). Only the samples with a 260/280 nm ratio between 1.8 and 2.1 and an RNA integrity number (RIN) higher than 8 were further processed. Total RNA samples (100 ng) were reverse-transcribed into double-stranded cDNA in presence of RNA poly-A controls, RNA Spike-In Kit, One-Colour (Agilent product number 5188-5282). The double-stranded cDNAs were *in vitro* transcribed in presence of Cy3-labelled nucleotides using a Low Input Quick Amp Labelling Kit, one-colour (Agilent product number 5190-2305). The Cy3-cDNA was purified using an ARNeasy mini kit, Qiagen (product number 74104 or 74106) and its quality and quantity was determined using NanoDrop ND 1000 and Bioanalyzer 2100. Only cDNA samples with a total cDNA yield higher than 2  $\mu\text{g}$  and a dye incorporation rate between 8  $\text{pmol } \mu\text{g}^{-1}$  and 20  $\text{pmol } \mu\text{g}^{-1}$  were considered for hybridization.

Cy3-labelled cRNA samples (1.65  $\mu\text{g}$ ) were mixed with a Agilent Blocking Solution, subsequently randomly fragmented to 100–200 bp at  $65^\circ\text{C}$  with Fragmentation Buffer, and resuspended in Hybridization Buffer using a Gene Expression Hybridization Kit (Agilent product number 5188-5242). Target cRNA Samples (100  $\mu\text{l}$ ) were hybridized to Whole Mouse Genome 4 $\times$ 44k OligoMicroarrays (Agilent G4122F) for 17 h at  $65^\circ\text{C}$ . Arrays were then washed using Agilent GE Wash Buffers 1 and 2 (Agilent product number 5188-5326), according to the manufacturer's instructions (One-Colour Microarray-Based Gene Expression Analysis Manual, <http://www.agilent.com>). An Agilent Microarray Scanner (Agilent product number G2565BA) was used to measure the fluorescent intensity emitted by the labelled target. The microarray data set has been uploaded to the NCBI Gene Expression Omnibus as record GSE43873, reorganized and filtered data can be downloaded in the Supplementary Information section (MicroarrayData.xls).

**Functional enrichment analysis.** The Gene Ontology term enrichment analysis was done with DAVID<sup>31,32</sup> on genes significantly more expressed (absolute  $\log_2$ (low/high crowding) gene expression value over 1.5) in FAK-expressing cells. Functional groups shown in the two networks have an enrichment value superior than 2 and are composed of at least 5 genes.

**Selection of candidate transcription factors.** The 19 transcription factors screened in the FAK-KO cells for their potential effect on *Abca1* mRNA expression were selected using a combination of three approaches. (1) Candidates have a binding site in all of the top 10 FAK suppressed genes defined with the microarray data. To perform this comparison, we used the Pscan algorithm (<http://www.beaconlab.it/pscan>) with the JASPAR database<sup>33</sup> (<http://jaspar.genereg.net/>). (2) Transcription factors having the strongest GO enrichment for lipid

homeostasis or (3) having a reported ChIP binding site or an effect on expression for ABCA1 in the literature (Supplementary Table 2).

**siRNA experiments.** All siRNAs were purchased from Qiagen. FAK-KO cells were cultured in 24-well plates, using standard conditions until reaching approximately 60% confluency (48–60 h) and transfected by forward transfection. Per well, 25 pmol samples of siRNA were mixed in 25  $\mu$ l of Opti-MEM and 0.5  $\mu$ l of Lipofectamine RNAiMAX were mixed with 24.5  $\mu$ l of Opti-MEM. After 5 min of incubation, solutions were mixed together and incubated for another 20 min at room temperature and transferred on the cultured cells for 60 h before RNA preparation.

**qPCR screening.** Silenced FAK-KO cells were washed with  $1\times$  PBS, RNA samples were prepared using NucleoSpinRNAII kit (Macherey Nagel), cDNA synthesis (Roche) was carried out with the Transcriptor High Fidelity cDNA Synthesis Kit (Roche) using poly-dT primers, in both cases following the manufacturer's protocol. Quantitative real-time PCR was performed in 384-well plates in an AB7900HT qPCR device (Applied Biosystems) using the following primers, forward ABCA1: 5'-CTGTAGACCTGGAGAGAAGCTTTC-3', reverse ABCA1: 5'-CAGCTCCA TGGACTTGTGATGAG-3' allowing amplification over the twelfth and thirteenth exons contained in all ABCA1 mRNA variants, and forward GAPDH: 5'-TCAAGGCTGAGAACGGGAAGCTTG-3', reverse GAPDH: 5'-AGCCTTCT CCATGGTGGTGAAGAC-3'. Relative mRNA amounts were calculated using GAPDH as an internal reference.

**Western blotting.** A431, FAK-WT and FAK-KO cells were cultured using standard conditions in 10-cm dishes. Low crowding cells were stopped after 2 to 2.5 days of growth, whereas high crowding cells were grown for 6 days (both including a final 12 h of serum starvation). Cells were washed with  $1\times$  PBS and disrupted in lysis buffer (0.5% sodium deoxycholate, 150 mM NaCl, 50 mM Tris-HCl, pH 7.2, 0.1% SDS, 1% Triton X-100, 0.2% NaN<sub>3</sub>), and 15  $\mu$ g of each protein extract was separated using 10% PAGE except for ABCA1 western blotting where 50  $\mu$ g of protein and 8% PAGE were used. Separated proteins were then transferred onto a membrane (Immobilon-P, 0.45  $\mu$ m, Millipore) using the humid chamber method. Transfer conditions are 80 mA overnight for ABCA1 western blotting, 250 mA for 90 min otherwise. Membranes were blocked with 4% BSA proteins in  $1\times$  TBS-T ( $1\times$  TBS, 0.1% Tween) for 1 h. Primary antibodies rabbit anti-pFAK (Cell Signaling no. 3283), rabbit anti-pPI(3)K (rabbit anti-pPI(3)K p85/p55 (T458/T199) antibody, Cell Signaling no. 4228), rabbit anti-pAKT ((T308) Cell Signaling no. 2965) were diluted at 1:1,000 and rabbit anti-actin (Cell Signaling no. 8456) at 1:5,000. Rabbit anti-TAL1 (Sc-12984, Santa Cruz) and rabbit anti-pFOXO3 (S253, no. 9466, Cell Signaling) were diluted at 1:200 and rabbit anti-ABCA1 (Abcam ab7360) at 1:500 in blocking buffer. HRP-conjugated secondary anti-mouse (no. 170-6516, BioRad) and anti-rabbit (no. 170-6515, BioRad) antibodies were diluted at 1:5,000 in the same buffer. Primary and secondary antibodies were applied overnight at 4°C and 60 min at room temperature, respectively. Signal was revealed with HRP substrate solution and imaged with a CCD camera (for antibody references see immunostaining section).

**ChIP experiments.** FAK-KO and FAK-WT cells were cultured using standard conditions in 10-cm dishes. Low crowding cells were stopped after 2 to 2.5 days of growth, whereas high crowding cells were grown for 6 days (both including a final 12 h of serum starvation). Experiments were carried out using the Chromatin Immunoprecipitation (ChIP) Assay Kit from Millipore following manufacturer's specifications except for the following changes. Fixation of cells was performed with 1.6 mM Di-thio bis-succinimidyl propionate (DSP) for 20 min, two short washes with  $1\times$  PBS at room temperature, and finally 1% paraformaldehyde for 20 min. 20  $\mu$ g of anti-TAL1 (Sc-12984, Santa Cruz), anti-FOXO3 (07-702, Millipore) and anti-LXR beta (Sc-34341, Santa Cruz) primary antibodies was added for 15 h at 4°C to the pre-cleared supernatant. Protein A beads were then added for 4 h. Reversion of crosslinking was done for 12 h at 55°C.

#### Lipid mass spectrometry

**Chemicals and lipid standards.** DLPC 12:0/12:0 (850335), PE 17:0/14:1 (PE31:1, LM-1104), PI 17:0/14:1 (PI31:1, LM-1504), PS 17:0/14:1 (PS31:1, LM-1304), C17:0 ceramide (860517), C12:0 SM (860583) and Glucosyl C8:0 Cer (860540) were used as internal lipid standards and were purchased from Avanti Polar Lipids Inc. (Alabaster, AL). Ergosterol was used as sterol standard and was purchased from Fluka (Buchs, Switzerland). Methyl tert-butyl ether (MTBE) was from Fluka (Buchs). Methyl amine (33% in absolute ethanol) was from Sigma Aldrich (Steinheim, Germany). HPLC-grade chloroform was purchased from Acros (Geel, Belgium), liquid chromatography-mass spectrometry (LC-MS) grade methanol and LC-MS grade ammonium acetate were from Fluka. LC-MS grade water was purchased from Biosolve.

**Cell culture.** FAK-WT cells were cultured using standard conditions in 10-cm dishes. Low crowding cells were stopped after 2.5–3 days of growth while high crowding cells were grown for 6 days (both including a final 12 h of serum

starvation). Cells were transfected with a human ABCA1-containing plasmid as described above or subjected to the transfection procedure without plasmid after one day of culture for low crowding cells or four days of culture for high crowding cells. Cells facing low or high crowding were collected two days after transfection. Cells were shortly washed with successively  $1\times$  PBS, 5% delipidated BSA, and three times with cold  $1\times$  PBS, scraped and pelleted at 800g for 5 min before lipid extraction.

**Lipid analysis.** Lipid extracts of 4 biological replicates of each of the 4 conditions (high crowding; high crowding + ABCA1; low crowding; low crowding + ABCA1) were prepared using the MTBE protocol<sup>34</sup> and measurements were made in 4 technical replicates, amounting to a total of 64 measurements at each mass spectrometer. Cell pellets were resuspended into 100  $\mu$ l of water and transferred into a 2 ml Eppendorf tube. Then 360  $\mu$ l methanol and a mix of internal standards were added (400 pmol DLPC, 1,000 pmol PE31:1, 1,000 pmol PI31:1, 3,300 pmol PS31:1, 2,500 pmol C12SM, 500 pmol C17Cer and 100 pmol C8GC). Samples were vortexed and 1.2 ml of MTBE was added. Samples were placed for 10 min on a multitube vortexer at 4°C (Lab-tek International) followed by an incubation for 1 h at room temperature on a shaker. Phase separation was induced by addition of 200  $\mu$ l MS-grade water. After 10 min of incubation at room temperature, samples were centrifuged at 1,000g for 10 min. The upper (organic) phase was transferred into a 13 mm glass tube with a Teflon-lined cap and the lower phase was re-extracted with 400  $\mu$ l artificial upper phase (MTBE/methanol/H<sub>2</sub>O 10:3:1.5). In total, 1,500  $\mu$ l of organic phase was recovered from each sample, split into three parts and dried in a CentriVap Vacuum Concentrator (Labconco). One part was treated by alkaline hydrolysis to enrich for sphingolipids and the other two aliquots were used for glycerophospholipid/phosphorus assay and sterol analysis, respectively. Glycerophospholipids were deacylated according to the method by Clarke & Dawson<sup>35</sup>. Briefly, 1 ml freshly prepared monomethylamine reagent (methylamine/H<sub>2</sub>O/n-butanol/methanol at 5:3:1:4 (vol/vol)) was added to the dried lipid extract and then incubated at 53°C for 1 h in a water bath. Lipids were cooled to room temperature and then dried. For desalting, the dried lipid extract was resuspended in 300  $\mu$ l water-saturated n-butanol and then extracted with 150  $\mu$ l H<sub>2</sub>O. The organic phase was collected, and the aqueous phase was reextracted twice with 300  $\mu$ l water-saturated n-butanol. The organic phases were pooled and dried in a CentriVap Vacuum Concentrator.

**Sterols analysis by gas chromatography-mass spectrometry (GC-MS).** One-third of total lipid extract was resuspended in 500  $\mu$ l of MS-grade chloroform/methanol (1:1) solution and injected into a VARIAN CP-3800 gas chromatogram equipped with a Factor Four Capillary Column VF-5ms 15 mm  $\times$  0.32 mm i.d. DF = 100. Identification and quantification of sterol species were performed using a VARIAN 320MS as described in ref. 36.

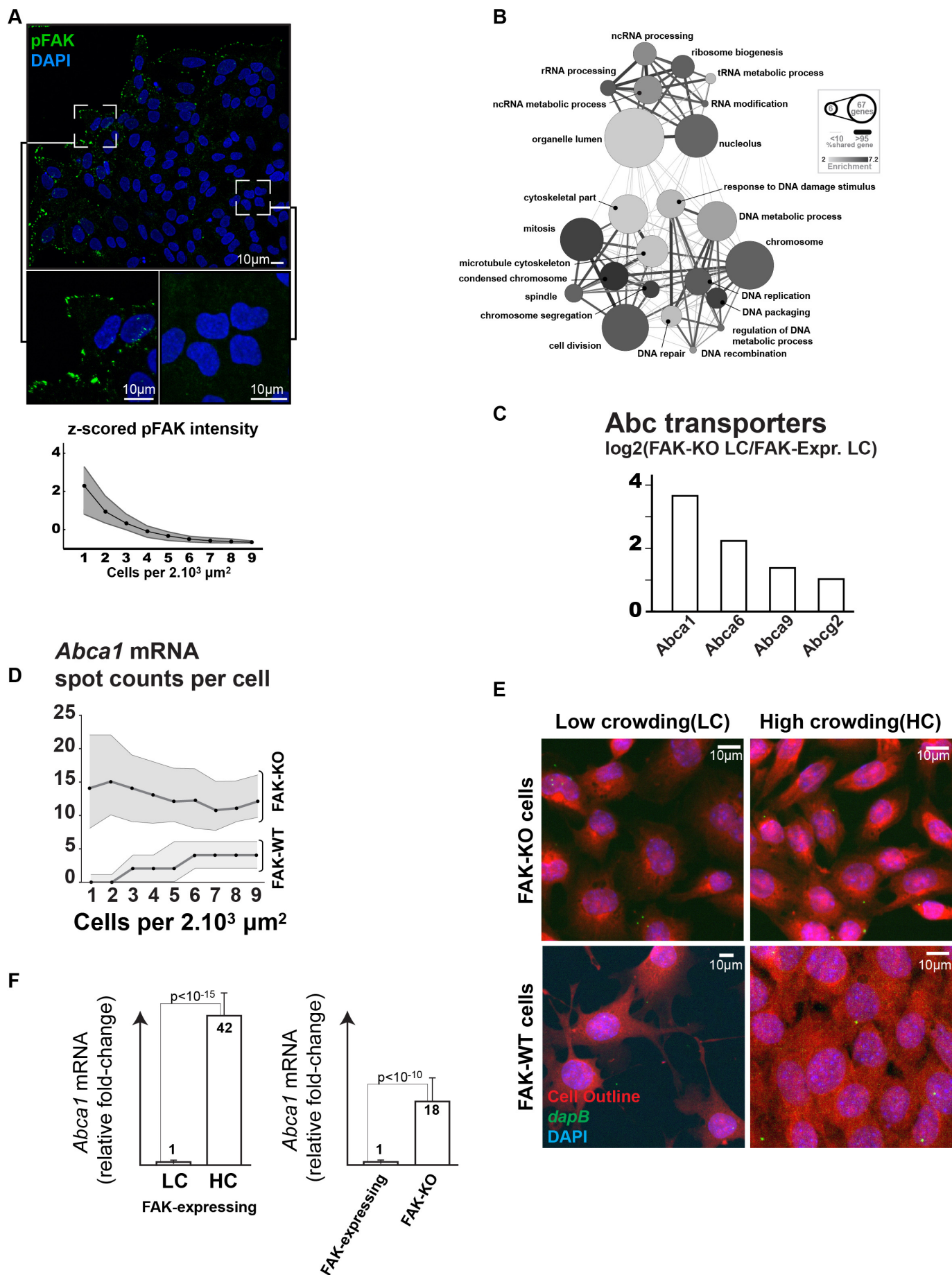
**Phospholipids and sphingolipids analysis by electrospray ionization mass spectrometry (ESI-MS).** Identification and quantification of phospholipid and sphingolipid molecular species were performed using multiple reaction monitoring with a TSQ Vantage Triple Stage Quadrupole Mass Spectrometer (Thermo Scientific) equipped with a robotic nanoflow ion source, Nanomate HD (Advion Biosciences). Each individual ion dissociation pathway was optimized with regard to collision energy. Lipid concentrations were calculated relative to the relevant internal standards as described in ref. 37 and then normalized to the total phosphorus content of each total lipid extract to adjust for difference in cell size, membrane content, and extraction efficiency.

**Determination of total phosphorus content.** The dried total lipid extract was resuspended in 250  $\mu$ l chloroform/methanol (1:1) and 50  $\mu$ l were placed into a 13 mm disposable pyrex tube. The solvent was completely evaporated and 0, 2, 5, 10, 20  $\mu$ l of a 3 mM KH<sub>2</sub>PO<sub>4</sub> standard solution were placed into separate pyrex tubes. To each tube 20  $\mu$ l of water and 140  $\mu$ l of 70% perchloric acid were added. Samples were heated at 180°C for 1 h in a hood. Tubes were then removed from the block and kept at room temperature for 5 min. Then 800  $\mu$ l of freshly prepared H<sub>2</sub>O/1.25% NH<sub>4</sub>Molybdate (100 mg/8 ml H<sub>2</sub>O)/10% ascorbic acid (100 mg/6 ml H<sub>2</sub>O) in the ratio of 5:2:1 were added. Tubes were heated at 100°C for 5 min with a marble on each tube to prevent evaporation. Tubes were cooled at room temperature for 5 min. 100  $\mu$ l of each sample was then transferred into a 96-well microplate and the absorbance at 820 nm was measured<sup>38</sup>.

27. Carpenter, A. E. *et al.* CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* **7**, R100 (2006).
28. Wippich, F. *et al.* Dual specificity kinase DYRK3 couples stress granule condensation/dissolution to mTORC1 signaling. *Cell* **152**, 791–805 (2013).
29. Rämö, P., Sacher, R., Snijder, B., Begemann, B. & Pelkmans, L. CellClassifier: supervised learning of cellular phenotypes. *Bioinformatics* **25**, 3028–3030 (2009).
30. Gaus, K., Zech, T. & Harder, T. Visualizing membrane microdomains by Laurdan 2-photon microscopy. *Mol. Membr. Biol.* **23**, 41–48 (2006).



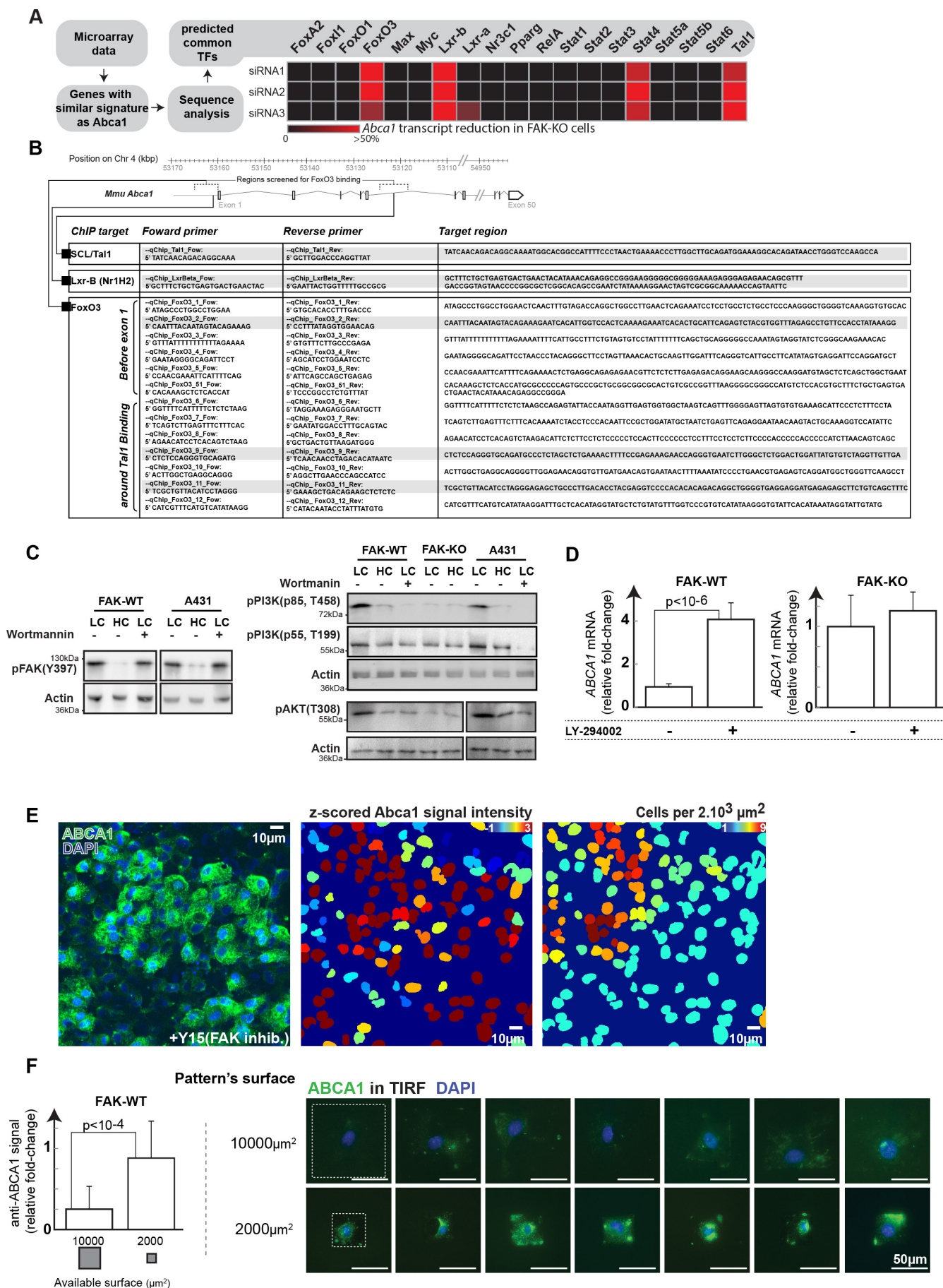
31. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4**, 44–57 (2009).
32. Huang, d. W., Sherman, B. T., Lempicki, R. & a.. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13 (2009).
33. Sandelin, A., Alkema, W., Engström, P., Wasserman, W. W. & Lenhard, B. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res.* **32**, D91–D94 (2004).
34. Matyash, V., Liebisch, G., Kurzchalia, T. V., Shevchenko, A. & Schwudke, D. Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics. *J. Lipid Res.* **49**, 1137–1146 (2008).
35. Clarke, N. G. & Dawson, R. M. *Alkaline O→N-transacylation*. 195, 301–306 (1981). A new method for the quantitative deacylation of phospholipids. *Biochem. J.* **195**, 301–306 (1981).
36. Guan, X. L., Riezman, I., Wenk, M. R. & Riezman, H. Yeast lipid analysis and quantification by mass spectrometry. *Methods Enzymol.* **470**, 369–391 (2010).
37. Epstein, S. *et al.* Activation of the unfolded protein response pathway causes ceramide accumulation in yeast and INS-1E insulinoma cells. *J. Lipid Res.* **53**, 412–420 (2012).
38. Rouser, G., Fleischer, S. & Yamamoto, A. Two dimensional thin layer chromatographic separation of polar lipids and determination of phospholipids by phosphorus analysis of spots. *Lipids* **5**, 494–496 (1970).



**Extended Data Figure 1 | Adaptation of the transcriptome to cellular crowding.** Related to Fig. 1. **a**, Immunofluorescence against phosphorylated FAK (Y397) in a population of A431 cells, corresponding curve shows single-cell phosphorylated FAK signals against local cell crowding (interquartile area is shown in grey, number of cells  $>10^4$ ). **b**, Gene Ontology enrichment network of genes that are induced by FAK in cells experiencing low crowding. Greyscale indicates enrichment, node-size number of genes, edge width between nodes number of overlapping genes. **c**, Histogram of ABC transporters more expressed in cells lacking FAK compared to cells expressing FAK when facing low crowding. **d**, Single-cell transcript counts of *Abca1* in  $1.2 \times 10^4$  FAK-KO and  $1.5 \times 10^4$  FAK-WT cells experiencing increasing levels of local crowding (interquartile area in grey). **e**, Control experiment of bDNA single-molecule

FISH against bacterial *dapB* transcripts in FAK-KO or FAK-WT cells experiencing low crowding or high crowding. Representative of  $10^4$  cells. **f**, Real-time PCR measurements of *Abca1* transcripts in cells at low and high local crowding in both FAK-expressing and FAK-KO cells in the presence of 10% FCS. Clearly, *Abca1* mRNA levels are much higher in FAK-expressing cells facing high crowding than in the same cells facing low crowding (s.d.,  $n = 4$  biological replicates each made of 3 technical replicates,  $P < 10^{-15}$ , *t*-test) but also in FAK-KO cells compared FAK-expressing cells (s.d.,  $n = 4$  biological replicates each made of 3 technical replicates,  $P < 10^{-10}$ , *t*-test). This indicates that FAK-dependent adaptation of *Abca1* transcription to cell crowding also operates in the presence of an abundant and homogeneous amount of growth factors and cytokines in the medium.



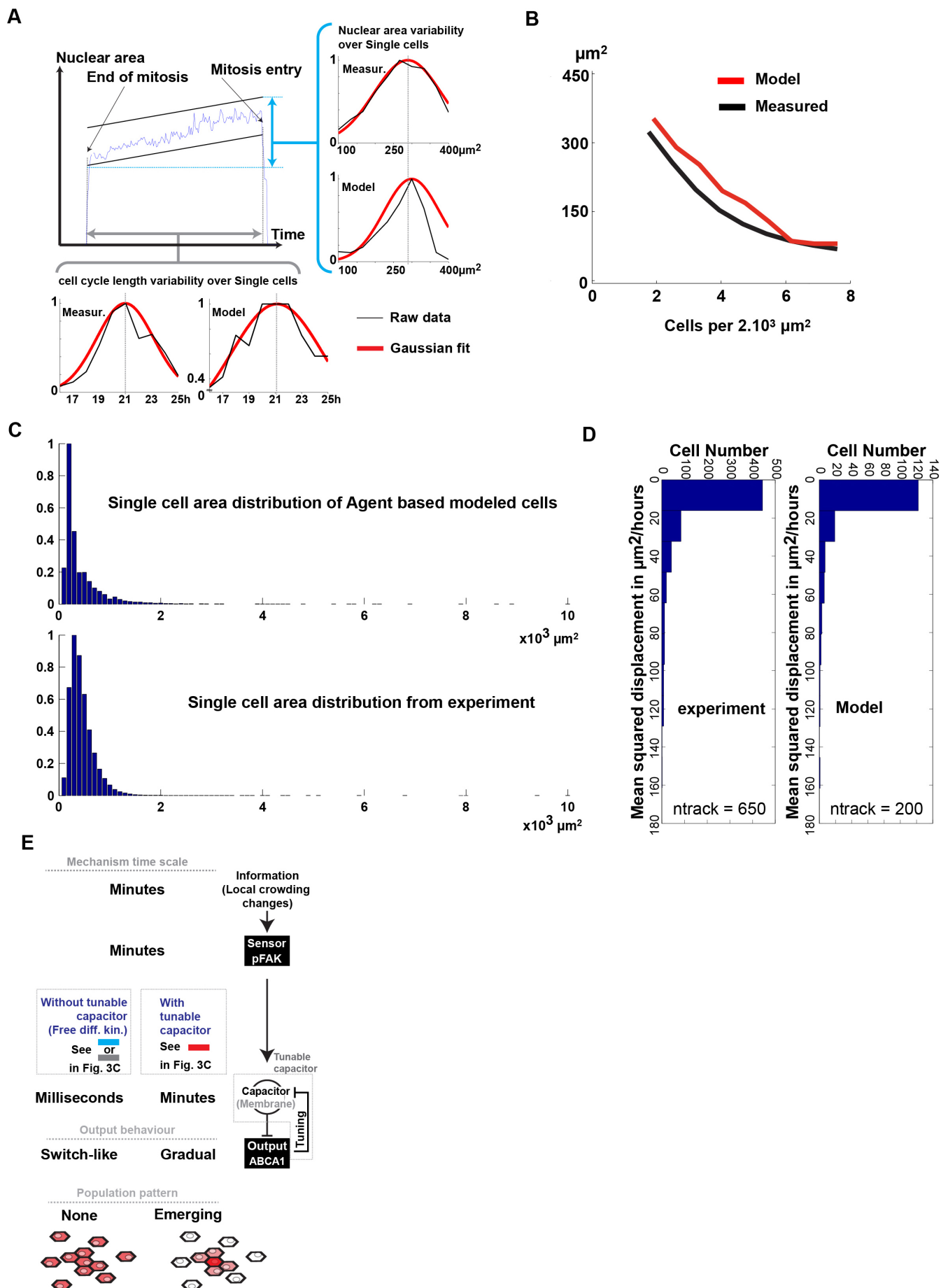


**Extended Data Figure 2 | FAK suppresses ABCA1 expression in cells at low crowding via TAL1 and FOXO3 in a cell-intrinsic way.** Related to Fig. 2.

**a**, Percentage reduction of *Abca1* mRNA in FAK-KO cells upon silencing of 19 potential transcription factors. **b**, Table of primers used for qRT-PCR amplification of *Abca1* DNA and corresponding genomic position. **c**, Western blots of pFAK, pPI(3)K and pAKT levels in FAK-WT and FAK-KO MEFs, and A431 cells at low crowding, high crowding or low crowding + wortmannin. **d**, Real-time PCR quantification of *Abca1* mRNA shows that treatment with LY-294002 alleviates the inhibitory effect of FAK on *Abca1* transcription in cells (at low crowding) expressing FAK (s.d.,  $n = 4$  biological replicates each made of 3 technical replicates,  $P < 10^{-6}$ ,  $t$ -test), whereas this treatment has no

significant effect on *Abca1* transcription in cells that lack FAK (s.d.,  $n = 4$  biological replicates each made of 3 technical replicates,  $P > 0.1$ ,  $t$ -test).

**e**, Immunofluorescence imaging of ABCA1 over a population of A431 cells in the presence of Y15 FAK inhibitor and related projection of single cell measurements onto nuclear segmentations. **f**, Quantifications of Abca1 protein expression in FAK-WT cells adhering to micropatterned surfaces of large (10,000  $\mu\text{m}^2$ ) or small (2,000  $\mu\text{m}^2$ ) area (<http://www.cytoo.com>) at long distance from potentially secreting neighbouring cells. This shows that space constraints are sufficient to trigger differences in Abca1 expression (s.d.,  $n = 100$  cells,  $P < 10^{-4}$ ,  $t$ -test).



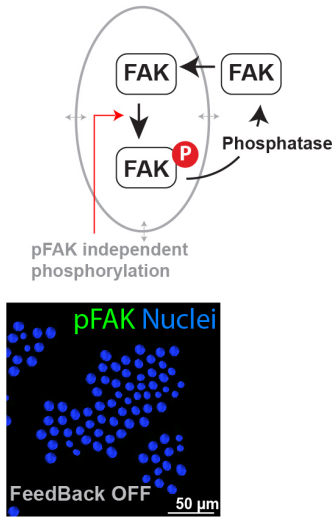


**Extended Data Figure 3 | Agent-based modelled single cells show**

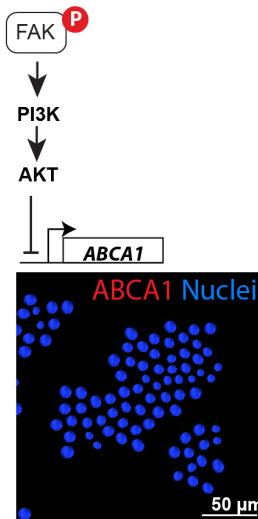
**characteristics similar to tracked cells.** **a**, Typical curve of the growth of the nucleus size of a single cell between two mitotic events (centre). Distribution of measured (number of tracks: 650) and agent-based modelled (number of tracks: 200) single-cell nucleus sizes (right histograms) and cell-cycle lengths (bottom histograms). Black, raw data, red, fitted Gaussian curve. Agent-based modelled cells and measured cells show similar distributions in cell-cycle length and nucleus size. **b**, Curve showing single-cell mean nuclear area against local cell crowding of measured (black, number of cells:  $>10^4$ ) and agent-based

modelled cells (red, number of cells:  $>10^3$ ). **c**, Histograms of single-cell area distribution of measured (number of cells:  $>10^4$ ) and agent-based modelled cells (number of cells:  $>10^3$ ) showing that distribution of emerging cell areas of modelled cells are matching those of measured cells even for extreme values. **d**, Histograms of single-cell mean square displacement distribution of measured (number of tracks: 650) and agent-based modelled cells (number of tracks: 200). **e**, Timescales of information sensing and processing steps in the FAK-ABCA1 system. Absence of a capacitor does not allow gradual patterns to emerge (switch-like behaviour).

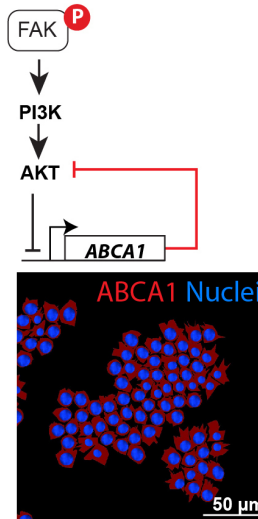
### A No auto-phosphorylation



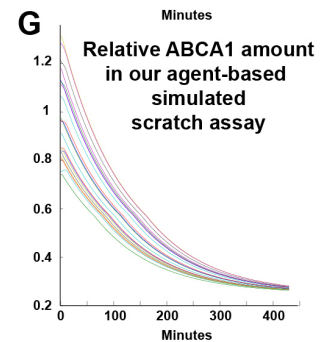
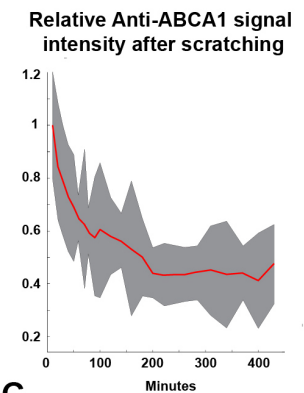
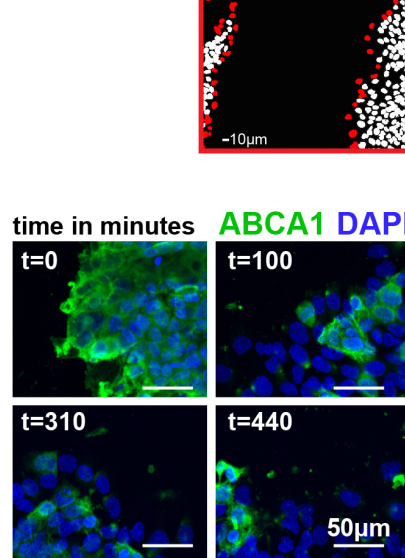
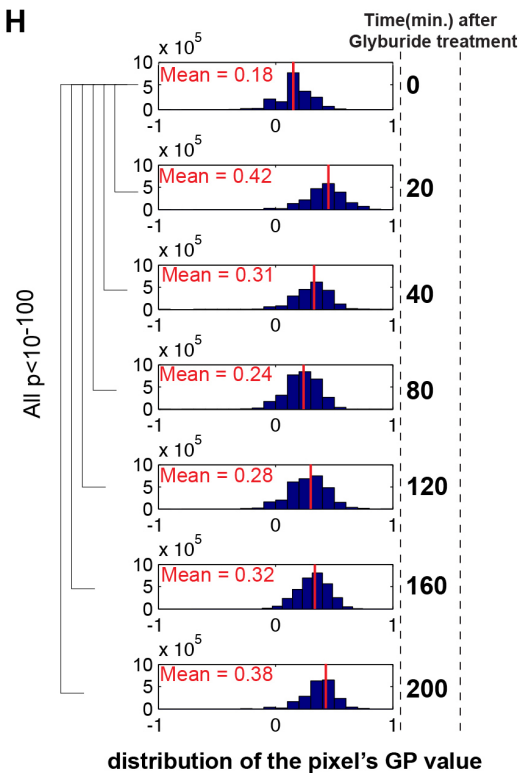
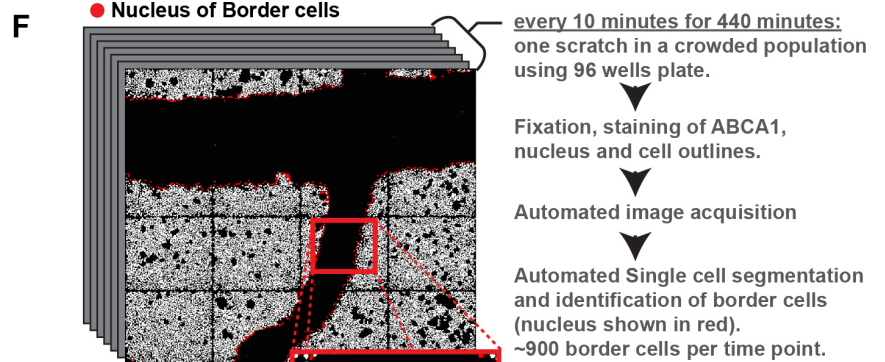
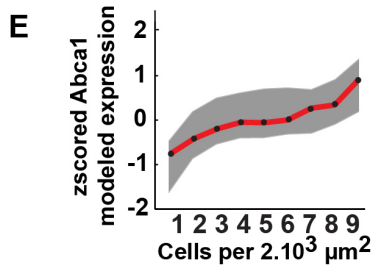
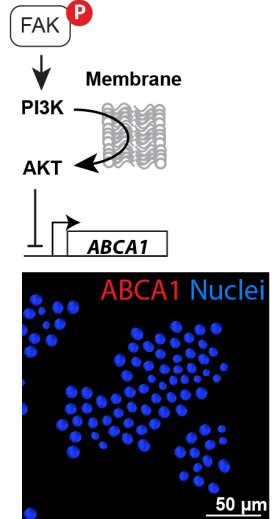
### B Free diffusion No feedback



### C Free diffusion direct feedback



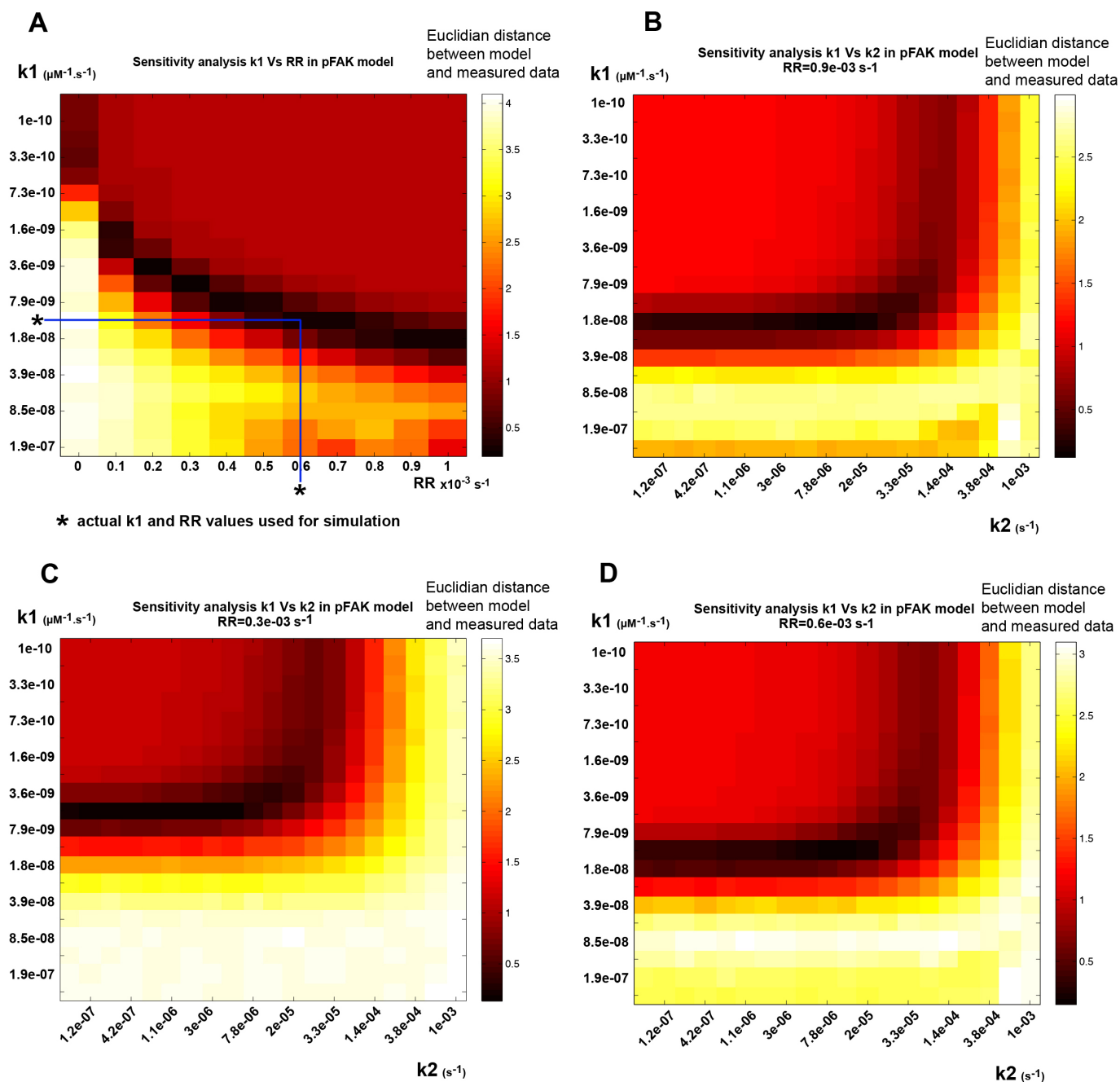
### D Membrane relay No feedback



**Extended Data Figure 4 | Alternative models do not lead to the emergence of gradual patterns in ABCA1 expression, and the full model recapitulates experimentally observed dynamics of reduction in ABCA1 expression in scratch assays.** Conclusions are parameter-independent, for details see mathematical appendix in the Supplementary Information. **a**, A FAK activation model without autophosphorylation does not result in a pFAK pattern in an agent-based modelled cell population. **b**, A FAK–ABCA1 model based on free diffusion of signalling molecules without or with **c**, addition of a putative direct inhibitory effect of ABCA1 on its own suppression does not result in a patterning of ABCA1 expression. **d**, Introduction of a membrane relay for AKT activation without ABCA1 feedback on the membrane relay does not result in a patterning of ABCA1 expression. **e**, Simulated single-cell ABCA1 variability over local crowding is similar to the variability seen in our experiments

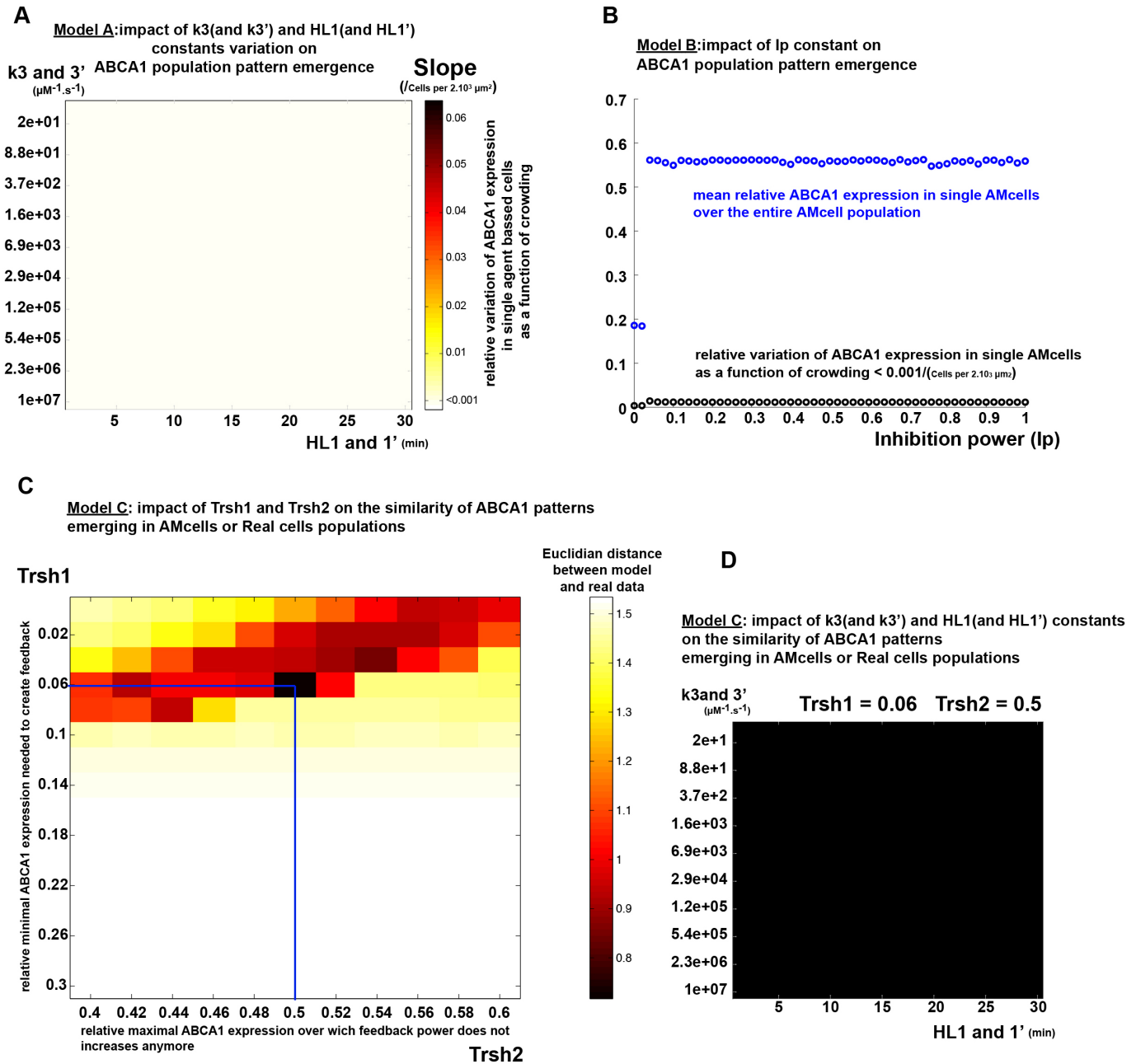
(see Fig. 2d). **f**, Scratch assays, at which cells at high crowding suddenly become exposed to free space to spread and followed over time, show that reduction of ABCA1 levels in these cells has a half-maximum effect at  $\sim 50$  min, and full effect at  $\sim 200$  min. **g**, This is in agreement with simulations of scratch assays using our cell-intrinsic Agent-based model of the FAK–ABCA1 system. The process was iterated thousands of times with random starting levels of ABCA1 similar to the variability seen in the experimental scratch assay. 20 representative curves are shown. In the simulations, it takes  $\sim 150$  min for the disappearance of half of ABCA1. **h**, Distributions of pixel GP values of FAK-KO cells stained with Laurdan at different time-points after treatment with glyburide. After just 20 min of drug treatment, the membranes of these cells become more ordered ( $P < 10^{-100}$ ,  $t$ -test, pixel distributions at each time point are made from  $2 \times 10^3$  cells).





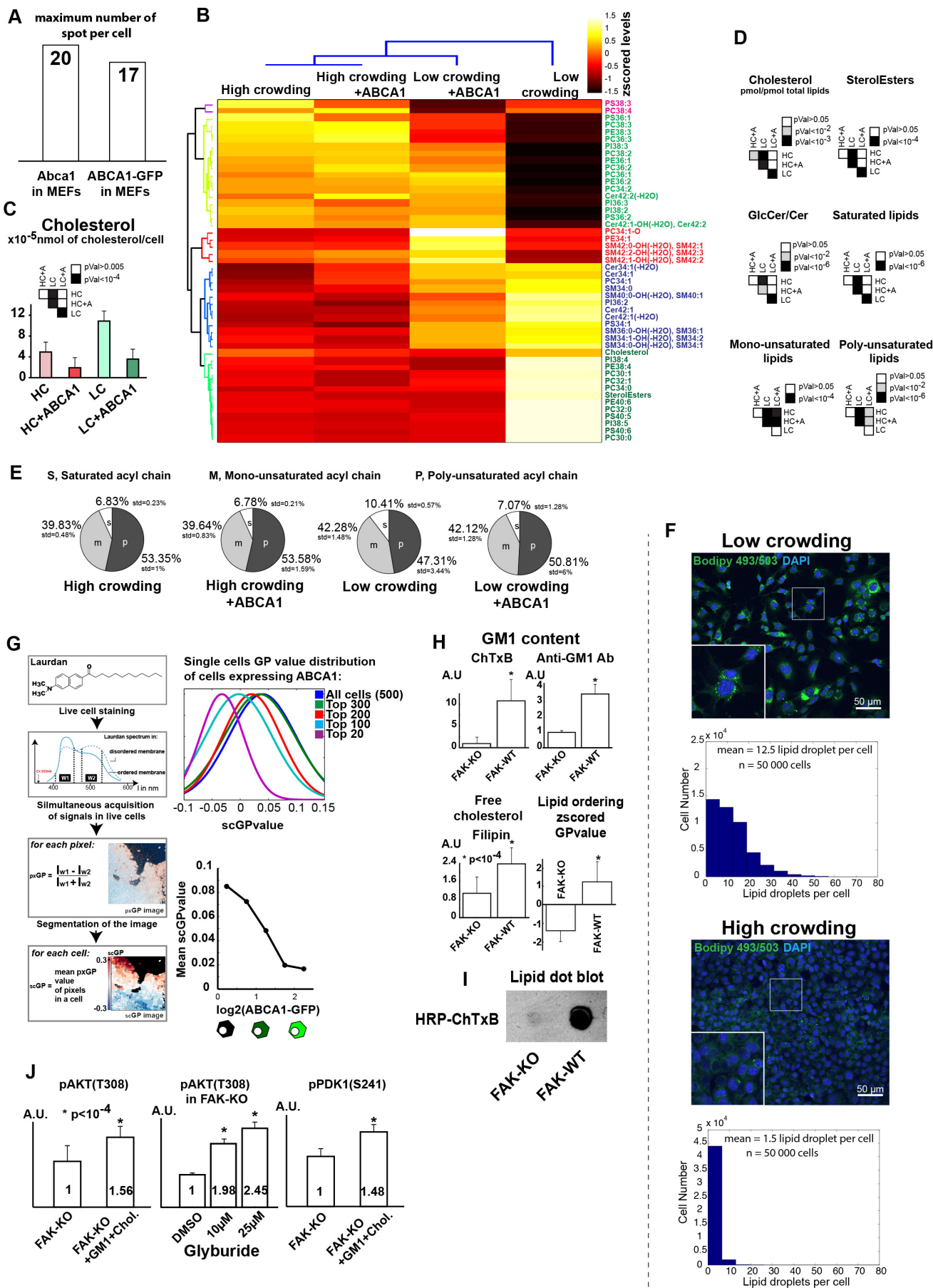
**Extended Data Figure 5 | Sensitivity analysis of the FAK activation model.**  
a, Heat map representing Euclidian distance between modelled and measured levels of pFAK in single cells as a function of local crowding when autophosphorylation constant  $k_1$  and removal rate RR varies. Stars represent the values used for further modelling; any pair of  $k_1$ -RR values with the same

low Euclidian distance will lead to the proper pFAK pattern. b-d, Same analysis for  $k_1$  and the FAK-independent phosphorylation of FAK rate  $k_2$  for a fixed RR value shows that FAK-independent phosphorylation of FAK has no effect on the formation of a pFAK pattern even if  $k_2$  is bigger than  $k_1$  by several orders of magnitude.



**Extended Data Figure 6 | Sensitivity analysis of the FAK to ABCA1 expression models.** **a**, Heat map representing the slope of ABCA1 expression against local cell crowding when  $k_3$  and  $3'$  and HL1 and  $1'$  vary over an extreme range of values for model A. This demonstrates that such topology cannot lead to emergence of gradual expression patterns ABCA1 expression as a function of local cell crowding. **b**, Mean relative ABCA1 expression in agent-based modelled cells as a function of its inhibition power ( $I_p$ ) in model B, where ABCA1 would be able to directly inhibit activation of AKT (or PI(3)K). This demonstrates that such direct feedback only leads to switch-like behaviour

where ABCA1 is either expressed or not in all cells of the population, independent of local cell crowding. Inhibition power represents the ABCA1 competitive inhibitory power. **c**, Heat map representing Euclidian distance between modelled and measured levels of ABCA1 in single cells as a function of local crowding when Trsh1 and Trsh2 vary in model C. **d**, The capacity of model C to generate a gradual expression pattern (low Euclidian distance is black) does not depend on  $k_3$  and  $3'$ , and HL1 and  $1'$ , demonstrating the central role of the membrane relay for gradual patterns to emerge.

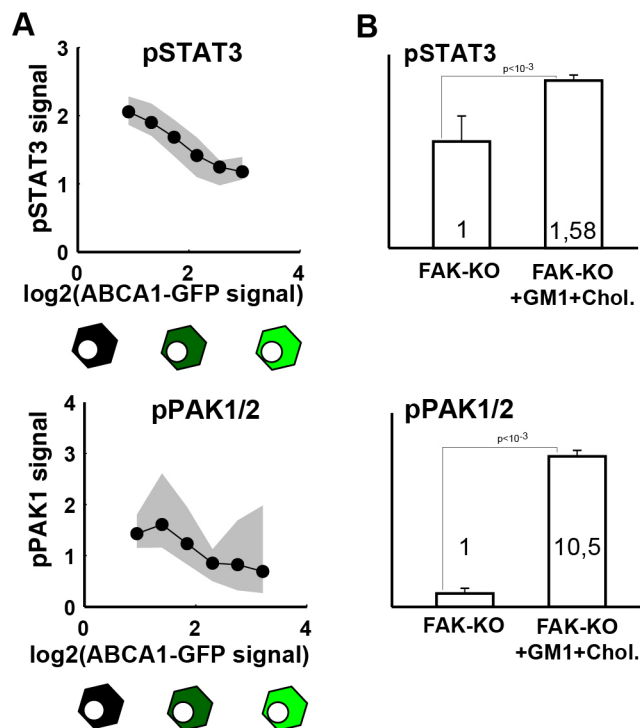




**Extended Data Figure 7 | The FAK–ABCA1 system adapts membrane lipid composition, ordering and signalling to local crowding.** Related to Fig. 4.

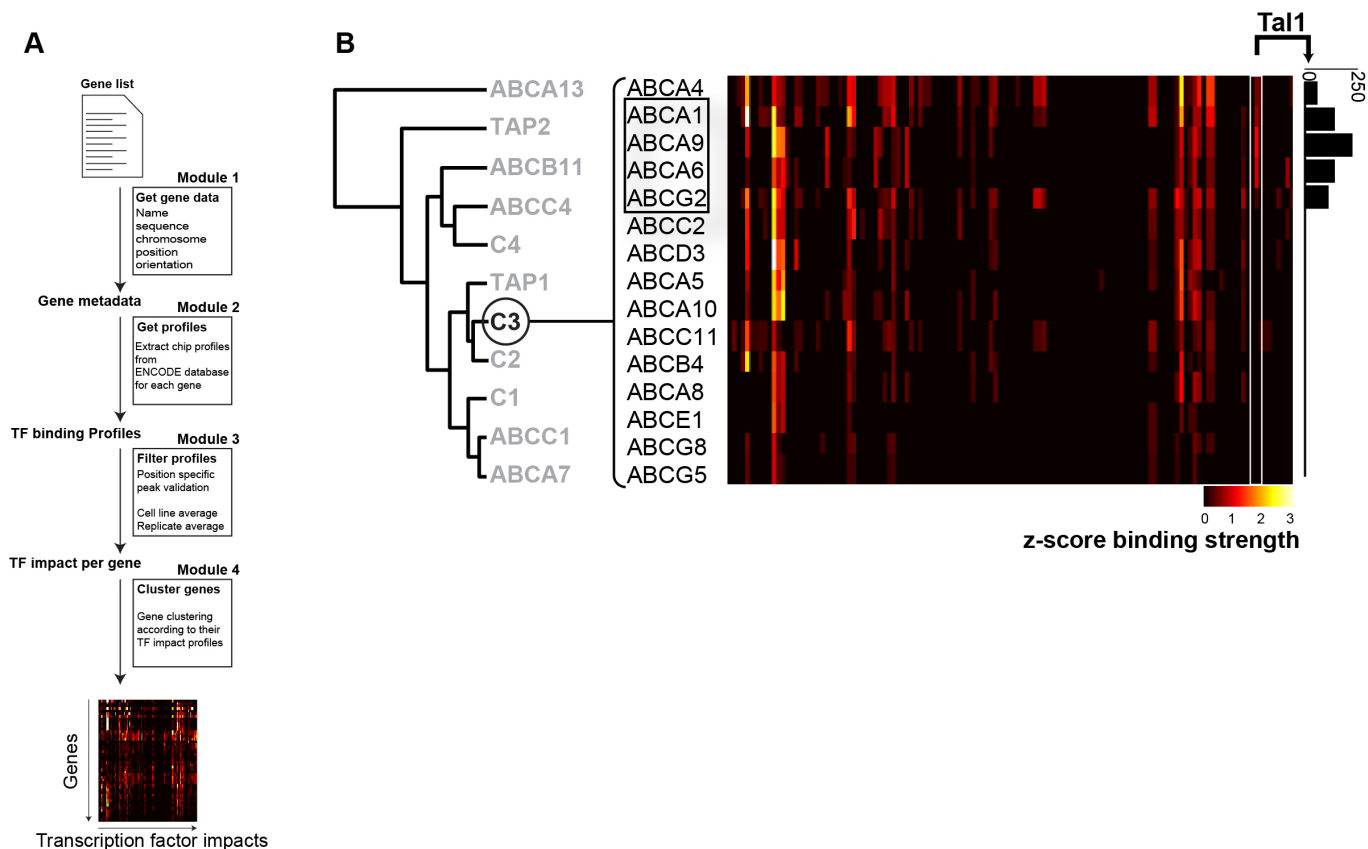
**a**, Histogram of transcript copy number (number of spots) per cell determined with bDNA single-molecule FISH against endogenous *Abca1* in cells at high crowding, or against *ABCA1–GFP* transcripts in cells at low crowding transfected with the pEGFP-N1-ABCA1 construct. This shows that plasmid-driven *ABCA1–GFP* expression in cells at low crowding does not exceed that of endogenous *Abca1* levels in cells at high crowding. **b**, Hierarchical clustering of lipid profiles of mouse embryonic fibroblasts grown at high crowding or low crowding conditions and transiently expressing ABCA1 from a plasmid (+ABCA1) or not. The clustergram shows the 48 lipid species that represent 80% of the total lipid amount. Colours correspond to pmol/pmol total lipid z-scored over the four conditions, colours of lipid names refer to their clusters. For complete lipid mass spectrometry data, see Supplementary Table 3. **c**, Histograms displaying the quantity of free cholesterol in nmol per cell ( $n = 4$  biological replicates, each the mean of 4 technical replicates, s.d.). **d**, *P* values related to the bar graphs in Fig. 4c. **e**, Pie charts representing the percentage of saturated, monounsaturated and polyunsaturated lipids for the four different conditions. **f**, Fluorescence imaging using Bodipy 493/503 dye of lipid droplets in low crowding ( $n = 5 \times 10^4$  single cells) or high crowding conditions

( $n = 5 \times 10^4$  single cells). This confirms that cells at low crowding contain a larger amount of cholesteryl-esters, which are stored in lipid droplets. **g**, Diagram summarizing the method to measure membrane ordering of a formaldehyde fixed population of cells at the single-cell level (left flow chart). Distributions of single-cell GP values for groups of cells that are the top 20, 100, 200, 300 ABCA1–GFP expressing cells compared to all cells (top right distributions,  $n = 500$  cells) and curve showing the relationship between single-cell ABCA1 expression and scGP value (bottom right curve,  $n = 500$  cells). **h**, Image-based quantification of free cholesterol (filipin), GM1 content (cholera toxin B binding or anti-GM1 antibody) and lipid ordering (Laurdan, as in panel **d**) in single MEFs with (FAK-WT) or without FAK (FAK-KO).  $n = 4$  experiments, each  $>10^4$  cells. \**P* values (*t*-test)  $< 10^{-4}$ . **i**, Because some GM1 may not be accessible in formaldehyde-fixed cells, we performed dot blot analysis of lipid extracts from FAK-KO and FAK-WT cells using HRP-conjugated cholera toxin B. This indicates that FAK-WT cells have higher levels of GM1 than FAK-KO cells. **j**, pAKT and pPDK1 immunostaining in cells without FAK (FAK-KO) exogenously loaded with GM1 and cholesterol (FAK-KO + GM1 + Chol.), treated with DMSO, or with 10 and 25  $\mu$ M glyburide in DMSO ( $n = 3$  experiments, each  $10^4$  cells, s.d., \**P* values (*t*-test)  $< 10^{-4}$ ).



**Extended Data Figure 8 | Phosphorylation of STAT3 and PAK1/2 are sensitive to ABCA1-mediated membrane perturbation.** **a**, Curve showing the relationship between ABCA1-GFP expression and phosphorylated STAT3 (T705) and PAK1/2 (T423/T402) amounts in single cells. **b**, Quantification of

immunostaining of phosphorylated STAT3 (T705) and PAK1/2 (T423/T402) amounts in FAK-KO cells after exogenous loading of the plasma membrane with cholesterol and GM1 (s.d.,  $n = 4$  experiments, each with  $10^4$  cells,  $t$ -test).



**Extended Data Figure 9 | Hierarchical clustering of human ABC transporters according to 118 transcription factor binding profiles from the ENCODE database.** **a**, Diagram of the algorithm used to generate ABC transporter clusters. **b**, Heat map of the cluster of ABC transporters containing

ABCA1, A9, A6 and G1 that share Tal1 binding (see bar graph representation of Tal1 binding on the right). These 4 ABC transporters are the same 4 ABC transporters that were found higher expressed in cells lacking FAK (FAK-KO) (see Extended Data Fig. 1c).



# Mechanical induction of the tumorigenic $\beta$ -catenin pathway by tumour growth pressure

María Elena Fernández-Sánchez<sup>1\*</sup>, Sandrine Barbier<sup>1\*</sup>, Joanne Whitehead<sup>1</sup>, Gaëlle Béalle<sup>2</sup>, Aude Michel<sup>2</sup>, Heldmuth Latorre-Ossa<sup>3</sup>, Colette Rey<sup>4</sup>, Laura Fouassier<sup>4</sup>, Audrey Claperon<sup>4</sup>, Laura Brullé<sup>5</sup>, Elodie Girard<sup>6</sup>, Nicolas Servant<sup>6</sup>, Thomas Rio-Frio<sup>7</sup>, Hélène Marie<sup>8</sup>, Sylviane Lesieur<sup>8</sup>, Chantal Housset<sup>4</sup>, Jean-Luc Gennisson<sup>3</sup>, Mickaël Tanter<sup>3</sup>, Christine Ménager<sup>2</sup>, Silvia Fre<sup>9</sup>, Sylvie Robine<sup>10</sup> & Emmanuel Farge<sup>1</sup>

**The tumour microenvironment may contribute to tumorigenesis owing to mechanical forces such as fibrotic stiffness or mechanical pressure caused by the expansion of hyper-proliferative cells<sup>1,2</sup>. Here we explore the contribution of the mechanical pressure exerted by tumour growth onto non-tumorous adjacent epithelium. In the early stage of mouse colon tumour development in the Notch<sup>+</sup>Apc<sup>+/-1638N</sup> mouse model, we observed mechanistic pressure stress in the non-tumorous epithelial cells caused by hyper-proliferative adjacent crypts overexpressing active Notch, which is associated with increased Ret and  $\beta$ -catenin signalling. We thus developed a method that allows the delivery of a defined mechanical pressure *in vivo*, by subcutaneously inserting a magnet close to the mouse colon. The implanted magnet generated a magnetic force on ultra-magnetic liposomes, stabilized in the mesenchymal cells of the connective tissue surrounding colonic crypts after intravenous injection. The magnetically induced pressure quantitatively mimicked the endogenous early tumour growth stress in the order of 1,200 Pa, without affecting tissue stiffness, as monitored by ultrasound strain imaging and shear wave elastography. The exertion of pressure mimicking that of tumour growth led to rapid Ret activation and downstream phosphorylation of  $\beta$ -catenin on Tyr654, impairing its interaction with the E-cadherin in adherens junctions, and which was followed by  $\beta$ -catenin nuclear translocation after 15 days. As a consequence, increased expression of  $\beta$ -catenin-target genes was observed at 1 month, together with crypt enlargement accompanying the formation of early tumorous aberrant crypt foci. Mechanical activation of the tumorigenic  $\beta$ -catenin pathway suggests unexplored modes of tumour propagation based on mechanical signalling pathways in healthy epithelial cells surrounding the tumour, which may contribute to tumour heterogeneity.**

To test the tumorous impact of early tumour growth pressure on non-tumorous epithelial tissues *in vivo*, apart from the mechanical stiffness characteristic of the microenvironment of late tumours<sup>1–6</sup> (see Supplementary Information 1), we first measured the strain deformation of live tumorous colonic crypts *ex vivo* under spinning disk microscopy. We used Notch/Apc mice, characterized by hyper-proliferative Notch overexpressing crypts within the Apc heterozygous genetic background (Notch<sup>+</sup>Apc<sup>+/-1638N</sup>)<sup>7</sup>. Colon crypt stress was deduced from their strain deformation induced by Notch<sup>+</sup> hyper-proliferative crypts, elasticity measurements, and Hooke's law. It was found to range from 0 to 2.4 kPa at 2 weeks ( $S = 1.2 \pm 1.2$  kPa

(mean  $\pm$  s.d.)), and from 0.4 to 5.6 kPa at 1 month ( $S = 3 \pm 2.6$  kPa) (Supplementary Information 2 and Extended Data Figure 1a–c). These values are consistent with the pathophysiological oncogene growth-induced stress of a few kPa induced by tumour growth<sup>2,8,9</sup>.

To apply and mimic a mechanical pressure in the same order of magnitude as the one generated by hyper-proliferative crypts *in vivo*, we developed a method based on the intravenous injection of stable ultra-magnetic liposomes (UML) encapsulating super-paramagnetic iron oxide nanocrystals. A 3-mm cylindric magnet was inserted subcutaneously in front of the colon, which caused loading of the colon tissue with the liposomes (Extended Data Figs 1d, 2a and Supplementary Information 3). Confocal microscopy showed UML loading, in the stromal cells surrounding the distal colon, one week after injection (Extended Data Fig. 2b). This remained stable, concentrated at  $250 \pm 50$  nmol of Fe(III) per g of tissue, during 1 month (Extended Data Fig. 2c). Immuofluorescence tissue analyses revealed UML internalisation within the mesenchymal cells surrounding the crypts, as determined by vimentin co-labelling (Fig. 1a).

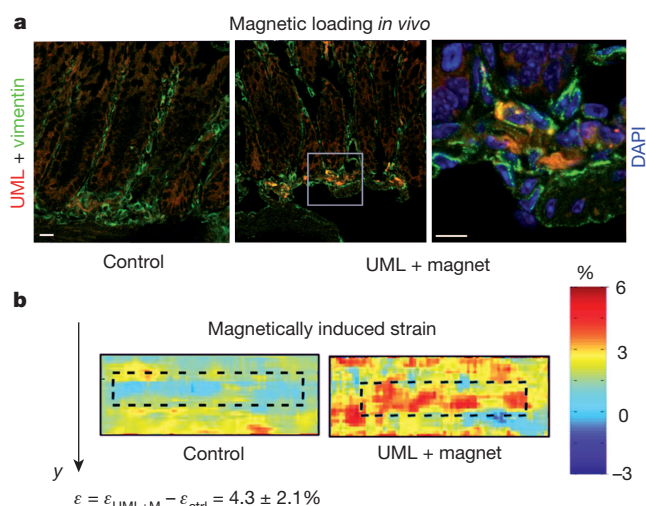
The next step was to measure the tissue stress associated with the presence of UML exposed to a magnetic field gradient in the colon of wild-type mice. The compression strain generated by the 3 mm (0.12 T) magnet on the UML-loaded tissue along the y direction of the wild-type colon was found to be  $\varepsilon = 4.3 \pm 2.1\%$  (Fig. 1b, Extended Data Fig. 2a, d–f and Supplementary Information 4). Quantitative Young's modulus maps of wild-type colons revealed a mean stiffness value of  $E = 30.1 \pm 3.5$  kPa (mean Young's modulus), which was not significantly changed by magnetic loading to  $E_{\text{UML}} = 35 \pm 3$  kPa (Extended Data Fig. 2g). These measurements enabled an estimation of a mean local uniaxial stress  $S = E \times \varepsilon = 1.194 \pm 0.610$  kPa generated by the magnet on the colon, making the assumption of an almost uniaxial strain (along y direction from the colon to the magnet) and consequently a one-dimensional Hooke's law ( $S = E \times \varepsilon$ ) approximation. This showed the existence of 100- $\mu$ m acoustic resolution domains of stress ranging from 0.6 to 1.8 kPa.

Notably, these values are consistent with the pathophysiological stress induced by tumour growth, as reported in the literature<sup>2,8,9</sup>. They are equivalent to the mechanical stress we measured in aberrant crypts generated in Notch/Apc mutant mice after 2 weeks to 1 month after aberrant crypt foci (ACF) development (Extended Data Fig. 1a).

We then applied this magnetically induced pressure *in vivo*, to the colon of Apc single-mutant mice (Apc<sup>+/-1638N</sup>), in which the  $\beta$ -catenin signalling pathway was specifically found to be mechanically activated

<sup>1</sup>Institut Curie, Centre de Recherche, PSL Research University, CNRS UMR 168, Physicochimie Curie Mechanics and Genetics of Embryonic and Tumour Development, INSERM, Fondation Pierre-Gilles de Gennes, F-75005 Paris, France. <sup>2</sup>UPMC, Sorbonne Universités, Laboratoire PHENIX Physico-chimie des Electrolytes et Nanosystèmes Interfaciaux, CNRS UMR 8234, F-75005 Paris, France. <sup>3</sup>Langevin Institut, Waves and Images ESPCI ParisTech, PSL Research University, CNRS UMR7587, Inserm U979, F-75005 Paris, France. <sup>4</sup>Sorbonne Universités, UPMC and INSERM, UMR-S 938, CDR Saint-Antoine, F-75012 Paris, France. <sup>5</sup>CNRS UMR3666/INSERM U1143, Endocytic Trafficking and Therapeutic Delivery, Institut Curie, Centre de Recherche, F-75005 Paris, France. <sup>6</sup>Bioinformatic platform, U900, Institut Curie, MINES ParisTech, F-75005 Paris, France. <sup>7</sup>Next-generation sequencing platform, Institut Curie, F-75005 Paris, France. <sup>8</sup>CNRS UMR 8612, Laboratoire Physico-Chimie des Systèmes Polyphasés, Institut Galien Paris-Sud, LabEx LERMIT, Faculté de Pharmacie, Université Paris-Sud, 92 296 Châtenay-Malabry, France. <sup>9</sup>CNRS UMR 3215/INSERM U934, Unité de Génétique et Biologie du Développement, Notch Signaling in Stem Cells and Tumors, Institut Curie, Centre de Recherche, F-75005 Paris, France. <sup>10</sup>CNRS UMR144, Compartimentation et dynamique cellulaires, Morphogenesis and Cell Signalling Institut Curie, Centre de Recherche, F-75005 Paris, France.

\*These authors contributed equally to this work.

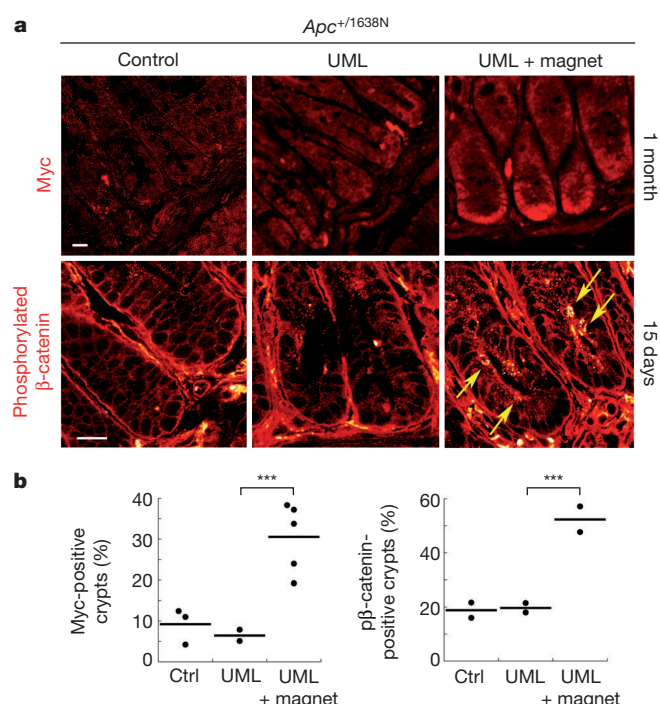


**Figure 1 | Stable magnetic loading of the colon tissue to mimic tumour growth pressure *in vivo*.** **a**, Rhodamine-labelled UML observed 1 week after injection in wild-type colon explants, in contrast to non-injected controls, localized in the connective tissue between the crypts. Co-localization of UML (in red) with vimentin (in green) (observed in  $n = 2$  mice, in the 200 crypts analysed). Scale bars are 10  $\mu\text{m}$ . **b**, Strain map images of wild-type colon 1 week after UML injection. Mean local strain compression of magnetized wild-type colon (measured in  $n = 6$  mice) compared to non-magnetized colon (measured in  $n = 3$  mice) is  $\varepsilon = 4.3 \pm 2.1\%$ .

in response to 20-min perturbation *ex vivo*<sup>10</sup>. This resulted in an oncogenic response leading to the increased expression of the  $\beta$ -catenin target tumorigenic *Myc*, *Axin2* and *Zeb1*<sup>10–12</sup> after 1 month of pressure application in the crypts (Fig. 2 and Extended Data Fig. 3a). Quantification by qRT-PCR (quantitative reverse-transcriptase PCR) revealed an increase in expression of *Myc*, *Axin2* and *Zeb1* by factors of 2.3, 2.5 and 3.5, respectively, due to magnetic pressure mimicking growth-induced mechanical tissue stress (Extended Data Fig. 3a).

Consistent with this transcriptional response, after 2 weeks of mechanical stress, we observed an increase of the phosphorylation of  $\beta$ -catenin on Y654 which prevents its interaction with junctional E-cadherin and allows its release from junctions<sup>13</sup> (Fig. 2, yellow arrows). Consistently, the phospho- $\beta$ -catenin did not co-localize with sub-apical junctional E-cadherins, but was detected more apically at the brush border, as assessed by co-localization with villin (Extended Data Fig. 4b). *Apc*-mutant crypts displayed an increased number of cells with a cytoplasmic enrichment in  $\beta$ -catenin, as well as nuclear enrichment, as shown by DNA co-staining with DAPI (Extended Data Fig. 3b, stained in white), attesting  $\beta$ -catenin accumulation both in the cytoplasm and nucleus in intestinal epithelial cells<sup>14</sup>. Consistent with the mechanical induction of *Myc* (Fig. 2), we observed an increase in the number of Ki67-positive proliferative cells (by 1.5-fold compared to control mice) after 2 weeks of mechanical stress (Extended Data Fig. 4a and Supplementary Note), and a consequent increase in the number of crypts larger than 1,500  $\mu\text{m}^2$  (Extended Data Fig. 3c). This corresponded to an increase in mean crypt size of 22% after 1 month, which is similar to the value recorded during the appearance of ACF in Notch/*Apc* mice, 1 month after Notch activation (Extended Data Fig. 3c).

Using small-animal colonoscopy, we could visualize the mechanical induction of ACF initiating 5 days after stress application, leading to adenoma development after 15 days in *Apc*<sup>+/-1638N</sup> mice (Extended Data Fig. 5a, b and Supplementary Information 5 with associated Extended Data Fig. 2c, h, i). Notably, no loss of *Apc* heterozygosity<sup>15</sup> could be detected in the epithelia in response to mechanical pressure at this time, in agreement with mechanically induced ACF formation (Extended Data Fig. 5c).

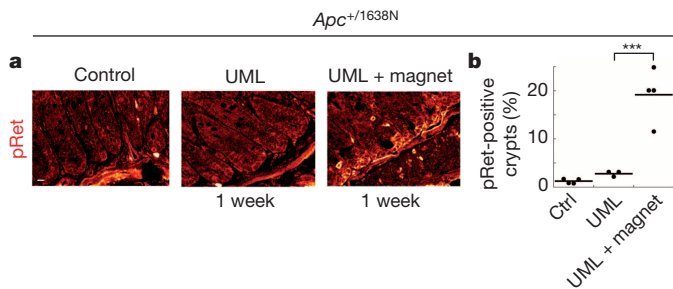


**Figure 2 | Activation of the Wnt/ $\beta$ -cat pathway by magnetic deformation mimicking tumour growth *in vivo*, in *Apc*-deficient mice colon.** **a**, Top, *Myc* expression in the crypts, 1 month after the initiation of mechanical strain with UML and magnet mimicking tumour strain (measured in  $n = 5$  mice, 487 crypts analysed), in control (measured in  $n = 3$  mice, 352 crypts analysed) and mice injected with UML without magnet (measured in  $n = 2$  mice, 167 crypts analysed), observed by immunofluorescence. Bottom,  $\beta$ -catenin Y654 phosphorylation after 15 days under UML with magnet conditions (measured in  $n = 2$  mice, 112 crypts analysed), compared to control (measured in  $n = 2$  mice, 129 crypts analysed) and to mice injected with UML without magnet (measured in  $n = 2$  mice, 92 crypts analysed). Scale bars are 10  $\mu\text{m}$ . **b**, Quantification of **a**.  $P > 0.05$  for all control versus UML comparisons. \*\*\* $P < 0.001$ , Student's *t*-test for comparison between all UML + magnet versus UML without magnet.

Therefore, a magnetic pressure mimicking the strain generated by tumour growth triggered the activation of the Wnt/ $\beta$ -catenin pathway, leading to the mechanical induction of *Myc*, *Axin2* and *Zeb1* expression, genes critically involved in tumour growth, and to ACF formation in *Apc*<sup>+/-1638N</sup> mice *in vivo*.

Then, we aimed to gain insights into the molecular mechanisms underlying the observed pro-tumorigenic effects of magnetically induced pressure. We found Ret mechanical phosphorylation as required upstream of the tumorigenic mechanosensitive pathway, *ex vivo* (Extended Data Figs 4d, e, g, 6a, b and Supplementary Information 6). To test whether tumour growth pressure could activate Ret *in vivo*, we first applied the magnetically induced pressure mimicking the  $\sigma = 1.194 \pm 0.660$  kPa tumour-growth-induced stress *in vivo*, to heterozygous *Apc*<sup>+/-1638N</sup> mutant mice. We observed Ret phosphorylation in *Apc*<sup>+/-1638N</sup> crypts at 1 week of induced mechanical stress (Fig. 3 and Extended Data Fig. 6c), initiating as early as after 30 min coincidental with UML accumulation (Extended Data Fig. 7a, b and Supplementary Information 6), consistent with the rapid dynamics associated with mechanotransductive processes and a direct mechanical activation of Ret. Ret activation, and downstream  $\beta$ -catenin Y654 phosphorylation and *Myc* induction, were all found to be independent of any indirect effect of UML loading per se (Figs. 2, 3 and Extended Data Fig. 7c, d), or long-range effects of liposomes or magnets either in the colon or outside the colon (Supplementary Information 7). We consistently found no expression of the Ret canonical ligands in the colon GDNF, artemin and neurturin (Extended Data



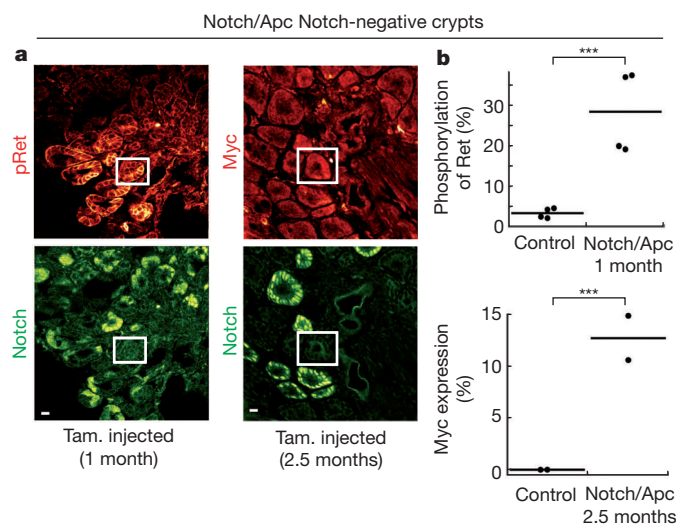


**Figure 3 | Mechanical activation of pY1062 Ret kinase after *in vivo* magnetic deformation of *Apc*-deficient mice colon.** **a**, Phosphorylated Ret Y1062 (pRet) in *Apc*<sup>+/1638N</sup> colon explants subjected to mechanical strain with UML and magnet mimicking tumour strain at 1 week (measured in *n* = 4 mice, 697 crypts analysed) compared to controls without UML (measured in *n* = 4 mice, 556 crypts analysed) and with UML without magnet (measured in *n* = 3 mice, 627 crypts analysed). Scale bar is 10  $\mu$ m. **b**, Quantification of **a**. *P* > 0.05 for control to UML comparison. \*\*\**P* < 0.001, Student's *t*-test.

Fig. 7e), which confirmed Ret mechanical activation in these experimental conditions.

We additionally found Akt-dependent Ret mechanical activation upstream of GSK-3 $\beta$  inactivation, a protein that prevents nuclear accumulation of the  $\beta$ -catenin released into the cytoplasm (Extended Data Fig. 7f, g and Supplementary Information 8)<sup>16,17</sup>.

Importantly, when we applied a magnetic pressure generating a tissue stress corresponding to *S* = 1.194  $\pm$  0.660 kPa to wild-type colon, we obtained the same results as in *Apc*<sup>+/1638N</sup> mice, namely Ret activation,  $\beta$ -catenin phosphorylation and release from sub-apical junctions and co-localized with villin and increased nuclear translocation of  $\beta$ -catenin (Extended Data Figs 4c and 6d, e). The observation of Myc overexpression and ACF formation with induced internal growth pressure were also observed, at 2 and 3 months, respectively (Extended Data Figs 6d, e, f and Supplementary Information 9), with no loss of the *Apc* allele in wild-type mice (Supplementary Information 9).



**Figure 4 | Mechanical activation of the Ret/ $\beta$ -catenin/Myc pathway in non-hyperproliferative Notch-negative crypts of Notch/*Apc* mice.**

**a**, Phosphorylation of Ret Y1062 (pRet) in Notch/*Apc* and overexpression of Myc in non-tumorous Notch-negative crypts (GFP negative, white boxes), compared to control. Tam., tamoxifen. Scale bars are 10  $\mu$ m. **b**, Quantitative results of **a**. \*\*\**P* < 0.001, Student's *t*-test. Phosphorylation of Ret Y1062: control (measured in *n* = 4 mice, 278 crypts analysed) and Notch/*Apc* (measured in *n* = 4 mice, 105 crypts analysed). Myc expression: control (measured in *n* = 2 mice, 200 crypts analysed) and Notch/*Apc* (measured in *n* = 2 mice, 150 crypts analysed). % indicates % of positive crypts.

Finally, we tested whether such a mechanotransduction pathway was activated by the endogenous physical pressure induced by hyperproliferative effects of Notch-activated crypts on Notch-negative crypts, in Notch/*Apc* mutant mice colon.

First, we found that both Ret and  $\beta$ -catenin were widely phosphorylated in aberrant Notch/*Apc* mutant colonic crypts. Enrichment in cytoplasmic and nuclear  $\beta$ -catenin were also observed 1 month upon Notch activation as well as Myc overexpression after 2.5 months (Extended Data Figs 8a, b, 4f and Supplementary Information 10).

Specifically, Ret phosphorylation,  $\beta$ -catenin phosphorylation and nuclear translocation, as well as Myc overexpression were observed not only in Notch-expressing crypts, but also frequently in non-tumorous crypts lacking Notch activation (Fig. 4 and Extended Data Fig. 8a, b). A substantial number of these crypts, which activates Ret, are surrounded by Notch negative crypts (Extended Data Fig. 9a). This is in agreement with long-range non cell-autonomous mechanical induction in healthy tissue by the strain developed by distant Notch hyperproliferative domains (Supplementary Information 11 with Extended Data Fig. 9a, b, c, d). No exonic mutations were detected in Ret in Notch/*Apc* mice epithelia or in UML + magnet *Apc*<sup>+/1638N</sup> mice, 3 months after tamoxifen injection and magnetic pressure, respectively, as well as in the wild-type copy of *Apc*<sup>+/1638N</sup> (not shown).

Altogether, our results show the mechanical activation of the tumorigenic Ret/ $\beta$ -catenin pathway in response to the pressure characteristic of hyperproliferative tumour growth pressure in colon epithelia in both *Apc*<sup>+/1638N</sup> and wild-type mice. We found that such a mechanotransduction process also takes place in non-tumorous *Apc*<sup>+/1638N</sup> heterozygous crypt cells compressed by Notch<sup>+</sup> hyperproliferative crypts of Notch/*Apc* mice (Supplementary Discussion 1). Based on these findings, we postulate that mechanical stimulation of tumorigenic pathways could potentially occur in normal tissues neighbouring any type of tumour, contributing to an unstable positive feedback loop between oncogene expression and tumour induction, enhancing the growth and spread of the tumour (Supplementary Discussions 1, 2, 3 and Extended Data Fig. 10 a–c).

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 1 August 2013; accepted 13 February 2015.

Published online 11 May 2015.

- Ghajar, C. M. & Bissell, M. J. Extracellular matrix control of mammary gland morphogenesis and tumorigenesis: insights from imaging. *Histochem. Cell Biol.* **130**, 1105–1118 (2008).
- Butcher, D. T., Alliston, T. & Weaver, V. M. A tense situation: forcing tumour progression. *Nat. Rev. Cancer* **9**, 108–122 (2009).
- Sawada, Y. *et al.* Force sensing by mechanical extension of the Src family kinase substrate p130Cas. *Cell* **127**, 1015–1026 (2006).
- Wozniak, M. A. & Chen, C. S. Mechano-transduction: a growing role for contractility. *Nature Rev. Mol. Cell Biol.* **10**, 34–43 (2009).
- Grashoff, C. *et al.* Measuring mechanical tension across vinculin reveals regulation of focal adhesion dynamics. *Nature* **466**, 263–266 (2010).
- Mammoto, A. *et al.* A mechanosensitive transcriptional mechanism that controls angiogenesis. *Nature* **457**, 1103–1108 (2009).
- Fre, S. *et al.* Notch and Wnt signals cooperatively control cell proliferation and tumorigenesis in the intestine. *Proc. Natl Acad. Sci. USA* **106**, 6309–6314 (2009).
- Delarue, M. *et al.* Mechanical control of cell flow in multicellular spheroids. *Phys. Rev. Lett.* **110**, 138103 (2013).
- Stilianopoulos, T. *et al.* Causes, consequences, and remedies for growth-induced solid stress in murine and human tumors. *Proc. Natl Acad. Sci. USA* **109**, 15101–15108 (2012).
- Whitehead, J. *et al.* Mechanical factors activate  $\beta$ -catenin-dependent oncogene expression in APC mouse colon. *HFSP J.* **2**, 286–294 (2008).
- Jho, E. H. *et al.* Wnt/ $\beta$ -catenin/Tcf signaling induces the transcription of Axin2, a negative regulator of the signaling pathway. *Mol. Cell. Biol.* **22**, 1172–1183 (2002).
- Kahlert, U. D. *et al.* Activation of canonical WNT/ $\beta$ -catenin signaling enhances *in vitro* motility of glioblastoma cells by activation of ZEB1 and other activators of epithelial-to-mesenchymal transition. *Cancer Lett.* **325**, 42–53 (2012).
- van Veelen, W. *et al.*  $\beta$ -catenin tyrosine 654 phosphorylation increases Wnt signalling and intestinal tumorigenesis. *Gut* **60**, 1204–1212 (2011).
- Clevers, H. Wnt/ $\beta$ -catenin signaling in development and disease. *Cell* **127**, 469–480 (2006).



15. Smits, R. *et al.* Loss of Apc and the entire chromosome 18 but absence of mutations at the Ras and Tp53 genes in intestinal tumors from Apc1638N, a mouse model for Apc-driven carcinogenesis. *Carcinogenesis* **18**, 321–327 (1997).
16. Stambolic, V. & Woodgett, J. R. Mitogen inactivation of glycogen synthase kinase-3  $\beta$  in intact cells via serine 9 phosphorylation. *Biochem. J.* **303**, 701–704 (1994).
17. Cohen, P. & Frame, S. The renaissance of GSK3. *Nature Rev. Mol. Cell Biol.* **2**, 769–776 (2001).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** This research was funded by the CNRS, INSERM, ARC (grant no.s 5030 and 29324), ANR (grant no.s 09PIRI0013-01 and 11 BSV5014-01), the Labex CellTisPhyBio (grant no. 11-LBX-0038), Fonds CSP, Fondation Pierre-Gille de Gennes, Fondation de France and the Inca (grant no. PLBIO13-172). High-throughput sequencing was supported by the grants ANR-10-EQPX-03 and ANR10-INBS-09-08 from the Agence Nationale de la Recherche (*investissements d'avenir*) and by the Canceropôle Ile-de-France. We thank the members of the Animal House Facility of Institut Curie, particularly Stéphanie Boissel, Virginie Dangles-Marie and Isabelle Grandjean, the IC Imaging facility members of PICT-IBiSA at BDD (UMR3215/U934) and Cell and Tissue Imaging Facility (UMR144 CNRS), Olivier Renaud's team and François Waharte, Thibaut Brunet for discussions on GSK3 phosphorylation, Pierre Gestraud for NGS analysis help, Jean-Michel Frapart and Sonia Lajnev for help in the

RPE experiments, and Didier Meseure (IC hospital) for anathomo-pathological analyses of the mechanically induced tumours.

**Author Contributions** M.E.F.-S. and S.B. designed and carried out experiments, interpreted data and prepared the manuscript. H.M. and S.L. produced the magnetic liposomes, G.B., A.M. and C.M. produced the ultra-magnetic liposomes, and M.E.F.-S. and S.B. contributed to their synthesis. M.E.F.-S., S.B. and C.R. performed the surgery of mice. L.F., A.C., M.E.F.-S., S.B. and C.H. set up the technique of implant the magnet and injected the mice. S.B., H.M. and S.L. measured, and defined respectively, the magnetic liposomes dosage administered intravenously to mice. M.E.F.-S., S.B., H.L., J.G. and M.T. performed acoustic experiments and interpreted the results. J.W. and M.E.F.-S. found the mechanical activation of pY1062 Ret kinase *ex vivo*. S.F. and S.R. provided all transgenic mice, designed and interpreted associated experiments. L.B. realized the colonoscopy with M.E.F.-S. and S.R. E.G., N.S. and T.R.-F. realized the sequencing and associated bioinformatics analysis. M.E.F.-S. realized RT-qPCR and all IF experiments except Myc IF *in vivo* and GSK3 IF, S.B. realized Myc IF *in vivo* and GSK3 IF and genotyping experiments and participated to the setting up of pRet IF. E.F. designed and coordinated the overall research and experiments, and wrote the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.F. ([efarge@curie.fr](mailto:efarge@curie.fr)).

## METHODS

Detailed protocols are deposited in Protocol Exchange, <http://dx.doi.org/10.1038/protex.2014.052>.

**Transgenic mice and tamoxifen administration.** All mice used in this work have been previously described:  $Apc^{+/1638N}$  (ref. 18),  $VilCreERT2/Nic$  and  $VilCreERT2/Nic;Apc^{+/1638N}$  (ref. 7), including males and females. C57Bl/6 inbred mice were obtained from Charles River Laboratories and females were used as wild-type controls.

For the experiments on  $VilCreERT2/Nic;Apc^{+/1638N}$  mice, four-week-old mice were injected intraperitoneally with tamoxifen (ICN) ( $50 \mu\text{g g}^{-1}$  of animal body weight) for five consecutive days. All animal experiments conformed to NIH guidelines for the care and use of animals and were approved by the local ethics committee (Approval no. 2012-0016). In this mouse model, a tamoxifen-inducible Cre recombinase (CreERT2) expressed under the control of the intestinal-specific villin promoter induces the expression of an intracellular and constitutively active form of the mitotic Notch 1 receptor (N1IC), along with a nuclear GFP reporter. As the  $VilCreERT2$  mice are mosaic, tamoxifen injection results in clones of mitotic cells expressing N1IC (N marked by GFP) within  $Apc^{+/1638N}$  heterozygous colonic cells.

**Synthesis of magnetic nanoparticles.** The aqueous suspension of magnetic nanoparticles was prepared using alkaline co-precipitation of  $\text{FeCl}_2$  (0.9 mol) and  $\text{FeCl}_3$  (1.5 mol) salts, according to Massart's procedure<sup>19</sup>. Superparamagnetic maghemite grains ( $\gamma\text{-Fe}_2\text{O}_3$ ) were obtained by oxidizing 1.3 mol of magnetite with 1.3 mol of iron nitrate (boiling solution). After magnetic decantation, 2 l of distilled water and 360 ml of  $\text{HNO}_3$  20% were added to the solution and the mixture was stirred for 10 min. Prepared maghemite nanoparticles were washed several times with acetone ( $3 \times 1$  l) and ether ( $2 \times 500$  ml) and suspended in water. Size sorting was performed by adding  $\text{HNO}_3$  (0.45 M) to the suspension followed by magnetic decantation. This operation was repeated with the deposit until a suitable particle size was obtained. Sodium citrate ( $n_{\text{Fe}}/n_{\text{Cit}} = 0.13$ , molar ratio) was added to the nanoparticles and the mixture was heated at  $80^\circ\text{C}$  for 30 min to promote adsorption of citrate anions onto their surface. Citrated nanoparticles were precipitated in acetone and suspended in water. The volume fraction and average size of the maghemite grains were determined by fitting the magnetization curve of nanoparticles using Langevin's Law. Particles of 7.7 nm (standard deviation  $\sigma = 0.37$ ) and of 9 nm diameter (standard deviation  $\sigma = 0.35$ , volume fraction of nanoparticles in the suspension  $\phi = 1.9\%$ , specific susceptibility  $\chi/\phi$  of 15.5) were obtained and used for magnetic liposomes and ultra-magnetic liposomes (UMLs), respectively. The aqueous medium was removed using a Macrosep advance centrifugal device (PALL) and nanoparticles were suspended again in a buffer (0.108 M NaCl, 0.02 M sodium citrate and 0.01 M HEPES, pH = 7.4).

**Preparation of ultra-magnetic liposomes (UMLs).** Solutions of 1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine (DPPC), 1,2-distearoyl-*sn*-glycero-3-phosphocholine (DSPC), 1,2-distearoyl-*sn*-glycero-3-phosphoethanolamine-*n*-[(carboxy(polyethyleneglycol)2000)](ammonium salt) (DSPE-PEG2000) and 1- $\alpha$ -phosphatidylethanolamine-*N*-(lissamine rhodamine sulfonyl B) (ammonium salt) (rhodamine-PE) in chloroform were purchased from Avanti Polar Lipids, Inc. UML were prepared by the reverse phase evaporation method established in ref. 20 and modified according to a previously described protocol<sup>21</sup>. In brief, a mixture of DPPC/DSPC/Rhod-PE/DSPE-PEG<sub>2000</sub> (84/10/1/5 mol%, 315  $\mu\text{l}$ ) was dissolved in 3 ml of diethyl ether (VWR) and 900  $\mu\text{l}$  of chloroform (Carlo Erba reagents). Thereafter, 1 ml of citrated magnetic nanoparticles dispersed in the buffer was introduced before sonication at room temperature for 20 min to produce a water-in-oil emulsion. Organic solvent was evaporated with a rotavapour R-210 (Buchi) at  $25^\circ\text{C}$  until the gel phase disappeared. Liposomes were filtrated through a 450 nm filter and purified from non-encapsulated magnetic nanoparticles by magnetic sorting using a strong magnet (Calamit Fe–Nd–B  $150 \times 100 \times 25$  mm). The operation was repeated three times every 2 h and liposomes (highly concentrated at 30% of nanoparticles in volume) were finally separated from the supernatant and recovered.

**Preparation of magnetic liposomes (MLs).** Magnetic liposomes were prepared according to a procedure already described<sup>22–26</sup>. In brief, the magnetic fluid used to load the liposomes consisted in an aqueous suspension of 8-nm nanocrystals of maghemite ( $\gamma\text{-Fe}_2\text{O}_3$ ) synthesized according to the procedure described above. In these conditions, superparamagnetic grains (from magnetization curve<sup>22,23</sup>) were produced. Final adjustment of both aqueous medium and maghemite concentration ( $5.4 \pm 0.1$  M Fe(III), checked by flame spectrometry) was performed by ultrafiltration (MACROSEP filter, cut-off 50 kD, Fisher Scientific Labosi) followed by addition of the same buffer as that used for liposome preparation (0.108 M NaCl, 0.02 M sodium citrate and 0.01 M HEPES, pH = 7.4). Final magnetic fluid was composed of nanoparticles of 20.3 nm in hydrodynamic diameter (0.1 poly-

dispersity index from QELS measurements) and stabilized by adsorption of citrate anions onto their surface.

Rhodamine-labelled magnetic-fluid-loaded liposomes (MFLs) were prepared by hydration of a thin lipid film (EPC:DSPE-PEG<sub>2000</sub>:rhodamine-PE; 94:5:1 mol%, determined by weight, precision of  $5.10^{-5}$  g) with equal volumes of magnetic fluid and buffer to get a total lipid concentration of 20 mM (checked by fluorescence spectroscopy). Hydration was followed by sequential extrusion under nitrogen pressure ( $<10$  bars) at  $25^\circ\text{C}$  through polycarbonate filters with decreasing pore diameters of 0.8  $\mu\text{m}$ /0.4  $\mu\text{m}$ /0.2  $\mu\text{m}$  (POREXICS, Osmotics, Livermore, USA). Liposomes of  $226 \pm 38$  nm in hydrodynamic diameter (from QELS measurements) were recovered. Non-entrapped maghemite particles were separated by gel exclusion chromatography (GEC) performed with a 0.4 cm  $\times$  5.8 cm Sephacryl S1000 superfine (Pharmacia) microcolumn (TERUMO 1 mL-syringe) beforehand saturated with EPC: DSPE-PEG<sub>2000</sub> (95:5 mol%) liposomes, prepared similarly to MFLs but without maghemite. The eluent was the buffer used for liposome preparation. Final Iron content of the preparation was checked by ESR analysis and found equal to  $44 \pm 1$  mM [Fe(III)], that is, 2.2 iron to lipid molar ratio (0.7% in volume concentration of nanoparticles).

**Quasi-elastic light scattering (QELS).** Hydrodynamic diameters were determined by using a photon correlator spectrometer (10 mW HeNe 632.8-nm laser, Zetasizer Nano ZS90, Malvern Instruments) at  $T = 25^\circ\text{C}$  and  $90^\circ$  scattering angle from the measured translational diffusion coefficients of the particles according to the Stokes–Einstein law for non-interacting spherical particles. Just before measurements, the samples were diluted with buffer to optimize the response of the apparatus. Unimodal distribution analysis was used and measurements were performed in triplicate.

**Injection of UMLs and MLs.** UMLs and MLs were injected intravenously in  $Apc^{+/1638N}$  and wild-type mice at 3–4 months (Extended Data Fig. 2a). The essential prerequisite of the experiments was to develop a methodology for introducing magnetic particles into the mesenchymal tissue of the colon at the intracellular level and at concentrations sufficiently high to allow subsequent magnetic manipulation. Furthermore, the mode of delivery of the magnetic material must preserve the tissue from any other stress, which would become competitive with the mechanic constraint generated magnetically. Thus the delivery must be realized through an indirect route of administration. Systemic delivery via intravenous injection performed outside the region of the colon ranks among the best ways provided that the pharmacokinetics of the magnetic particles is optimized to reduce first-pass hepatic clearance and permit observable distribution in the colon tissue. In this respect, we used submicron liposomes sterically stabilized by polyethylene glycol (PEG) coating as bioavailability-enhancing carrier of the magnetic particles. Indeed, PEGylated magnetic-fluid-loaded liposomes not exceeding 200 nm in diameter have reliably been proved to be long-circulating systems as intact vesicle structures without leakage of their inner content, therefore aptly averting dilution of the magnetic material<sup>22–24</sup>. Moreover they have shown to diffuse from the vasculature into the interstitial tissues without loss of structure integrity and at the intracellular level as well<sup>25,26</sup>. Said otherwise, the containment of the magnetic particles required for magnetic manipulation and beforehand adjusted within the liposomes is totally conserved during their passage through the vascular endothelium towards the surrounding tissues and upon cellular uptake.

**Mechanical deformations.** Ex vivo global compression. To apply an ex vivo mechanical compression colon samples were treated as previously described with a confinement box of 3 mm thickness for 0.8 kPa pressure applied<sup>10</sup>. For kinase inhibition experiments, PP1, PP2 (30  $\mu\text{M}$ ; BioSource), SU6656 (20  $\mu\text{M}$ ; Sigma-Aldrich), sunitinib (20  $\mu\text{M}$ ; Sigma-Aldrich), vandetanib (10  $\mu\text{M}$ ; Selleck Chemicals), ponatinib (100 nM; Selleck Chemicals), or an equivalent volume of the vehicle DMSO (diluted to 1 in 500), was added to reduced serum medium 20 min prior starting the compression.

**In vivo magnetic deformation.** Mice were anaesthetized with isoflurane (delivered at 2% for maintenance and 1.5% for induction in oxygen) and a 3 mm diameter strong magnetic field gradient magnet positioned subcutaneously on the back of the mice in front of the colon (Extended Data Fig. 2a). The fluorescent UML (associated with rhodamine-PE phospholipids) were diluted in a 10 mM HEPES pH 7.4, 20 mM  $\text{Na}_2\text{Citrate}$ , 108 mM NaCl buffer solution to a final concentration of 0.1 M and injected in the lateral caudal veins of the tail. The only rare side-effect of injection was a small hematoma that disappeared in two to three days later. Magnet implantation caused only local skin inflammation that was treated with antiseptics and disappeared in three to four days. No effect on intestinal transit was detected nor dysfunction of the colon, which was largely isolated from skin.

**Pharmacological inhibitors.** The Src family inhibitors used in this study are PP1 and PP2<sup>27</sup>, and SU6656 as a specific inhibitor of Src<sup>28</sup>. The inhibitors of Ret used in this study are sunitinib<sup>29</sup>, vandetanib<sup>30</sup> and ponatinib<sup>31,32</sup>.

**Live tissue imaging and analysis.** The distal colon was dissected from adult mice, rinsed with PBS and incubated at 37 °C in Leibovitz's L-15 medium + Glutamax-1 supplemented with 10% fetal bovine serum and 40 µg ml<sup>-1</sup> gentamicin. Individual tissue segments were opened longitudinally and placed villi up between two coverslips. Images were obtained with an inverted Olympus IX70-Ropper spinning disc microscope coupled to a Cool Snap HQ2 camera.

**Iron concentration analyses.** Iron concentration analyses were realized by electron spin resonance (ESR) analysis. UML and magnetic liposome biodistribution was assessed by determining the amount of iron oxide in excised tissues using ESR spectrometry. ESR spectra were acquired using an Elexsys E500 ESR spectrometer (Bruker Biospin) operated at: resonant frequency ≈ 9.2 GHz; microwave power = 10.11 mW, receiver gain = 60 db and magnetic field modulation amplitude = 10G, at room temperature. After the sacrifice of the mouse, distal colon samples were lyophilized for 24 h. Maghemite nanocrystal suspensions used for each liposome sample and of known iron concentration were used as standards to construct a calibration curve ( $R^2 = 0.99$ ) for quantitative purposes. ESR spectra were collected as the first derivative of the absorbed microwave power versus the magnetic field. A double integration was applied to obtain the area under the curve of the absorption-field curve, which is proportional to the number of resonating electronic spins in the measured sample. All tissue samples studied were corrected for background tissue absorption of the microwave radiation using tissue samples of control mice not exposed to UML. All the double integrations were calculated over the field interval (1500–5500 Gauss). Iron amounts were normalized by tissue sample weights (mol of Fe per mg tissue). Iron concentration in  $Apc^{+/1638N}$  at 1 week was confirmed to be on the same order of magnitude than at 1 week in the wild type before being quantitatively evaluated for each time point conditions based on the quantitative comparison of the mean density of fluorescence associated to UML between each  $Apc^{+/1638N}$  sample, and the mean fluorescence value of the wild-type samples (using ImageJ), based on ten immunohistological samples per mouse.

**Acoustic analysis.** A distal colon explant was completely embedded in an agar-gelatin phantom. The ultrasonic probe was fixed on one side of the phantom and a small magnet approached axially towards the colon by a support.

To measure the axial deformation (along  $y$  direction), strain imaging is performed, by acquiring an ultrasonic image before and after the magnet was displaced. The magnetic-force-induced tissue strain, which was measured by estimating local tissue displacement along the  $y$  direction from the first acquisition to the second one. Note that the time between moving the magnet and acquiring the data is less than 5 s, which is short enough to avoid creep behaviour. Average values were obtained from at least four measurements for each condition.

Shear wave elastography (SWE) was performed to image quantitative tissue stiffness. A remote and tiny palpation (tens of µm) is induced by the acoustic radiation force of an initial focused ultrasonic beam and this radiation force generates a shear wave propagating along the  $x$  direction. Tissue displacements induced by this shear wave are mapped by performing ultrafast ultrasonic imaging. The local shear wave speed is linked to stiffness.

Dissected colon samples were embedded in agar-gelatin phantoms (2% agar with 5% gelatin)<sup>33</sup>. Ultrasound images, so called B-mode and elasticity images were acquired using a high frequency ultrasound probe (15 MHz, 256 elements, Vermon) driven by an ultrafast imaging device (Aixplorer, Supersonic Imagine)<sup>34</sup>. SWE, that is, Young's modulus ( $E$ ) quantitative imaging, was performed by using the supersonic shear wave imaging (SSI) technique *ex vivo* and *in vivo*<sup>35,36</sup>.

The SSI technique is based on the ultrafast ultrasound imaging of a shear wave induced by the radiation force of an initial focused ultrasonic beam acting as a remote palpation in tissues (Extended Data Fig. 2e). Under the assumptions of a local constant density  $\rho$  and a locally incompressible and isotropic elastic medium, the propagation speed  $v_s$  of the tracked shear wave is directly linked to the Young's modulus  $E$  (in kPa) characterizing the local stiffness via the relationship:  $E = 3\rho \times v_s^2$ .

The assumption of constant density of *in vivo* soft biological tissues here is valid. Indeed, in biological soft tissues, the density is almost constant  $\rho \approx 1,000 \text{ kg m}^{-3}$  due to the very high water content of soft tissues. However, small density variations exist as shown in extensive past studies<sup>37,38</sup>. From these studies, the mean density (among connective tissues, muscle, fat, blood cells, plasma, cornea, spinal cord, spleen, testis) is  $1,052 \pm 47 \text{ kg m}^{-3}$ . Thus, the normalized standard deviation in soft tissues is 4.7%.

Also, due to their high fluid content, many soft biological tissues and gels exhibit nearly incompressible behaviour under physiological loading; they are constrained to undergo essentially volume-preserving deformations and motions. Thus, it is commonly accepted in the field of tissue elasticity measurements that the Poisson's ratio of tissue has a value between 0.49 and 0.4999, meaning that tissue is nearly incompressible<sup>39,40</sup>. Of course, this incompressible behaviour is only ensured provided that the conditions of interest do not allow the water to diffuse into or out of

the tissue during the period of interest. This is the case in the SWE approach, due to the very small (micrometric) displacements induced by the shear wave used to probe local elasticity. Such tiny displacements do not induce water diffusion outside of the organ.

Although the assumption of local isotropy has to be done to derive the Young's modulus from the shear wave speed measurements, it is not possible currently to prove its validity *in vivo*. In particular tissue such as muscles, an elastic anisotropy has already been proven *in vivo*<sup>41,42</sup>. However, the *in vivo* assessment of such anisotropic elasticity was only made possible in particular configurations, such as the human biceps, because the structural organization and orientation of the muscular fibre bundles is highly identical over a large region of interest. Other tissues like breast, arteries, liver are today assumed as isotropic in the field of elastography. Even if a local anisotropy could potentially exist in these tissues, we postulate that it should remain quite small. Indeed, it was recently shown that the anisotropy of shear modulus in the *in vivo* kidney was quite small with a fractional anisotropy of cortex and medulla <20% (see Fig. 3. of ref. 43) despite a more important tissue organization in the kidney (due to the alignment of the pyramids) than in the other tissues.

Extensive calibration experiments were performed in the past to demonstrate the ability of SWE to quantify the Young's modulus of tissues. The standard deviation of Young's modulus quantification was demonstrated to be lower than 5% on calibrated phantoms mimicking biological tissue properties<sup>44,45</sup>. A small magnet (3 mm in diameter) was axially approached towards the colon by steps of 0.5 mm until completing 3 mm of absolute axial displacement. For each position of the magnet, strain images ( $\varepsilon$ ) were calculated by comparing raw frequency ultrasound images acquired at two consecutive steps<sup>46</sup>. Cumulative one-dimensional strain along the  $y$  direction was obtained by summing all strain images (Extended Data Fig. 2f). Although the magnet could induce some stress in the full volume in the three directions of space, it is here considered that the strain in the lateral ( $x$ ) and elevational ( $z$ ) directions remain small compared to the measured axial ( $y$ ) strain. Under this assumption of a force that is in majority in the axial  $y$  direction, the quantitative stress  $S$  applied by the magnetic field acting on ferrofluids trapped in colon tissues was retrieved by calculating the one-dimensional Hooke's law:  $S = E \times \varepsilon$  (ref. 47). Using Hooke's law, the standard deviation of  $S$  was experimentally estimated and found to be equal to 0.61 kPa. It is in good agreement with a 0.74 kPa standard deviation of  $S$  derived from equation (1) that describes the influence of  $E$  and  $\varepsilon$  uncertainties ( $35.0 \pm 3.0 \text{ kPa}$  and  $4.3 \pm 2.1\%$ , respectively) on the uncertainty of  $S$ :

$$\text{std}(S) = \sqrt{\text{var}(E \times \varepsilon)} = \sqrt{\text{std}(E)^2 \times \text{std}(\varepsilon)^2 + \text{std}(E)^2 \times \varepsilon^2 + \text{std}(\varepsilon)^2 \times E^2} \quad (1)$$

During the experiments, collection was performed no more than 4 s after the compression induced by the magnet motion. Such a small delay ensures that one can avoid any creep behaviour. Indeed, the relaxation time for typical human tissues under compression is of the order of several tens to hundreds of seconds<sup>48</sup>.

The distance between the colon and the magnet was evaluated using ultrasound imaging *in vivo*, with anal injection of a water gel transparent to ultrasound in the colon lumen to visualize the colon.

**Western blot.** Isolation of epithelial cells of the colon was performed by incubation of tissue explants in 3 mM EDTA and 0.5 mM DTT in phosphate-buffered saline (PBS) for 45 min. Cell pellets were washed in PBS and lysed in radioimmune precipitation assay buffer (50 mM Tris-Cl (pH 8.0), 150 mM NaCl, 1% Nonidet P-40, 0.5% deoxycholate and 0.05% SDS) plus 1% phosphatase inhibitor cocktail, and 1% protease inhibitors. The protein content of the supernatant was measured by the colorimetric reaction RD DC protein assay (Bio-Rad). An equivalent quantity of protein (20 µg) was resolved by SDS-PAGE, transferred to a nitrocellulose membrane (Invitrogen), and hybridized with the appropriate antibodies, followed by detection using enhanced chemiluminescence (Pierce ECL Plus Western Blotting Substrate, Thermo Scientific). Protein levels were quantified using ImageJ. The following antibodies were used: anti-pY1062 Ret (1:100, Santa Cruz Biotechnology), anti-GAPDH (1:5,000, abcam), and horseradish peroxidase (HRP)-conjugated secondary antibodies (1:1,000, Jackson ImmunoResearch).

**RNA isolation and qRT-PCR.** Total RNA isolation was performed by using NucleoSpin RNAII (Macherey-Nagel), according to the manufacturer's instructions and quantified with a NanoDrop ND-1000 spectrophotometer. First-strand cDNA was synthesized by using AccuScript High Fidelity 1st Strand cDNA Synthesis kit (Agilent Technologies). For qPCR, reactions were run on a real-time PCR system (ABI Prism 7900; Applied Biosystems) and gene expression was detected with Power SYBR Green (Applied Biosystems). Relative gene expression was determined by normalizing to reference genes  $\beta 2$ -microglobulin and TATA box binding protein TFIID (transcription factor IID), using the comparative threshold cycle ( $C_T$ ) method<sup>49</sup>. Results were expressed as fold change of each sample versus control. The primers used in each reaction were as follows: *Myc* forward 5'-TCAAGAGGCGAACACACAAC-3' and reverse 5'-GGCCTTTT



CATTGTTTTCCA-3'; *Axin2* forward 5'-CGCCACCAAGACCTACATACG-3' and reverse 5'-ACATGACCGAGCCGATCTGT-3'; *Zeb1* forward 5'-TGGCAAGACAACGTGAAAGA-3' and reverse 5'-AACTGGGAAATGCATCTGG-3'; *TFII* forward 5'-CCACGGACAAGTGGCT-3' and reverse 5'-GGCTCATAGCTACTGA-3';  $\beta$ 2-microglobulin forward 5'-GCTATCCAGAAAA CCCCTCAA-3' and reverse 5'-AGGCGGGTGGAACTGTGT-3'. Statistical data analyses were performed using a two-tailed unpaired Student's *t*-test between any two groups ( $n = 7$ ).

**Accession numbers for gene expression data.** The UniProt accession numbers for sequences used in the gene expression analyses are: *Axin2*, O88566; *Myc*, P01108; and *Zeb1*, Q64318.

**Next-generation sequencing.** Targeted re-sequencing of *Apc* and *Ret* gene panel was designed with Ion amplicon designer v3.2. Multiplex PCR primers were selected to amplify the coding sequences of the selected genes: *Apc* (98.39% of the gene coding sequence covered by the design), *Ctmb1* (100%), *Ret* (99.8%), *Gsk3b* (100%) and *Axin2* (96.26%) with an exon padding of 25 bp. Sequencing libraries were built with the Ion Amplicon Library Kit v2.0 following manufacturer's recommendations (Life Technologies). In brief, 10 ng of each DNA sample was mixed with designed primer pools and supplied PCR mix to amplify targeted genomic regions. PCR products corresponding to PCR primers were partially digested and barcoded Ion adapters were ligated at the 5' and 3' extremities of digested PCR products. After purification of ligation products, PCR amplification followed by quality control was performed. Generated sequencing libraries were then pooled in equimolar ratio and sequencing templates were prepared on an Ion One Touch system with the Ion OneTouch 200 Template Kit v2 DL (Life Technologies). After selective enrichment and loading on a Ion 318v2 chip, positive sequencing templates were sequenced on a Ion Torrent PGM with the Ion PGM Sequencing Kit 200 v2 (Life Technologies). Three technical replicates for each mouse DNA sample were processed in parallel to determine the technical variability to accurately evaluate the copy number variation on the targeted genes. **Bioinformatics analysis.** Sequenced reads were aligned on the *mm10* reference genome using the TMAP software (v3.6.2, Life Technologies). Around 95% of the reads were aligned on the targeted genes with a mean depth of coverage of 1750 $\times$ . The variant calling was performed using the Ion Torrent Variant Caller (v3.6.2, Life Technologies) and annotated with the RefSeq genes and dbSNP138 databases using the Annovar software<sup>50</sup>. Variants found with a low coverage ( $\leq 30\times$ ) or with a strand bias were filtered out.

**Analysis of *Apc* status by genotyping.** Isolation of epithelial cells of the colon was performed by incubation of tissue explants in 3 mM EDTA and 0.5 mM DTT in PBS for 45 min. The supernatant was then discarded, 10 ml of cold PBS added, and the mixture was shaken vigorously by hand before being spun for 20 min at 153g at 4 °C to obtain the cell pellet. Total DNA isolation was performed by using QIAamp DNA mini kit (Qiagen), according to the manufacturer's instructions. For PCR, gene expression was detected with BioTaq Red DNA Polymerase (Abcys). For *Apc* genotyping, the following primers were used: primer PA2, TCAGCCATG CCAACAAAGTCA, primer PN3, GCCAGCTCATTCTCCACTC and primer C2, GGAAAGTTTATAGGT. Cycling conditions were 5 min at 94 °C (1 cycle); 30 s at 94 °C, 45 s at 59 °C, and 345 s at 72 °C (35 cycles); and 5 min at 72 °C (1 cycle). The presence of the wild-type allele is indicated by a 300-bp PCR fragment (primers PA2 and C2) and the mutant allele by a 400-bp PCR fragment (primers PA2 and PN3).

**Analysis of *Apc* status by DNA sequencing.** Isolation of epithelial cells of the colon was performed by incubation of tissue explants in 3 mM EDTA and 0.5 mM DTT in phosphate-buffered saline (PBS) for 45 min. The supernatant was then discarded, 10 ml of cold PBS added, and the mixture was shaken vigorously by hand before being spun for 20 min at 153g. at 4 °C to obtain the cell pellet. DNA was isolated according to the QIAamp DNA mini kit (Qiagen) in order to be analysed by the Sequencing Platform of the Institut Curie. A dedicated sequencing panel targeting the coding sequence of the *Apc* and *Ret* genes was used. Only one variant (D2086G) was detected in the *Apc* gene in one control mouse and one mouse subject to mechanical pressure. This single nucleotide change is referenced as a common polymorphism in the dbSNP database (rs47505115). No Y1062 mutation was detected in the *Ret* gene. No exonic mutation was detected in the *Ret* gene (not shown).

**Histology and immunohistochemistry.** Freshly dissected colon samples were treated as previously described<sup>10</sup>. The commercial antibodies used were: anti-Myc N262 (1:50, Santa Cruz Biotechnology), Twist H81 (1:50, Santa Cruz Biotechnology),  $\beta$ -catenin (1:100, BD), pY654  $\beta$ -catenin (1:50, abcam), pY1062 *Ret* (1:50, Santa Cruz Biotechnology), *Zeb1* (1:50, Santa Cruz Biotechnology), vimentin (1:300, Sigma), Ki67 (1:200, abcam), pS9 GSK-3 (1:50, Cell Signalling), pY568/570 Kit (1:100, Santa Cruz Biotechnology), pY412 Abl (1:50, Santa Cruz Biotechnology), pY209/211 Hck (1:50, Sigma), pY537 Yes (1:100, Santa Cruz Biotechnology), pY1238/1239 Ron (1:50, Santa Cruz Biotechnology), pSer473

Akt (1:50, Santa Cruz Biotechnology), E-cadherin (1:100, Santa Cruz Biotechnology), villin (1:100, Santa Cruz Biotechnology), GDNF (1:50, abcam), neurturin (1:200, abcam), artemin (1:200, abcam), anti-mouse Cy3 (1:500, Jackson ImmunoResearch), anti-rabbit Alexa 568 (1:600, Molecular Probes), anti-rabbit Alexa 350 (1:200, Molecular Probes) and anti-rabbit Alexa 488 (1:200, Molecular Probes). All images were taken with a Zeiss LSM710 microscope at the Platform for Cell and Tissue Imaging of the Institut Curie.

Nuclear  $\beta$ -cat translocation was evidenced using the co-localization highlighter function of the co-localization analysis ImageJ plugin ([http://www.uhnresearch.ca/facilities/wcif/imagej/colour\\_analysis.htm](http://www.uhnresearch.ca/facilities/wcif/imagej/colour_analysis.htm)). Co-localization was coded in a binary way as white pixels (positive for co-localization) and quantified by measuring the density of white pixels, with control parameters adjusted at the threshold showing no co-localization in controls.

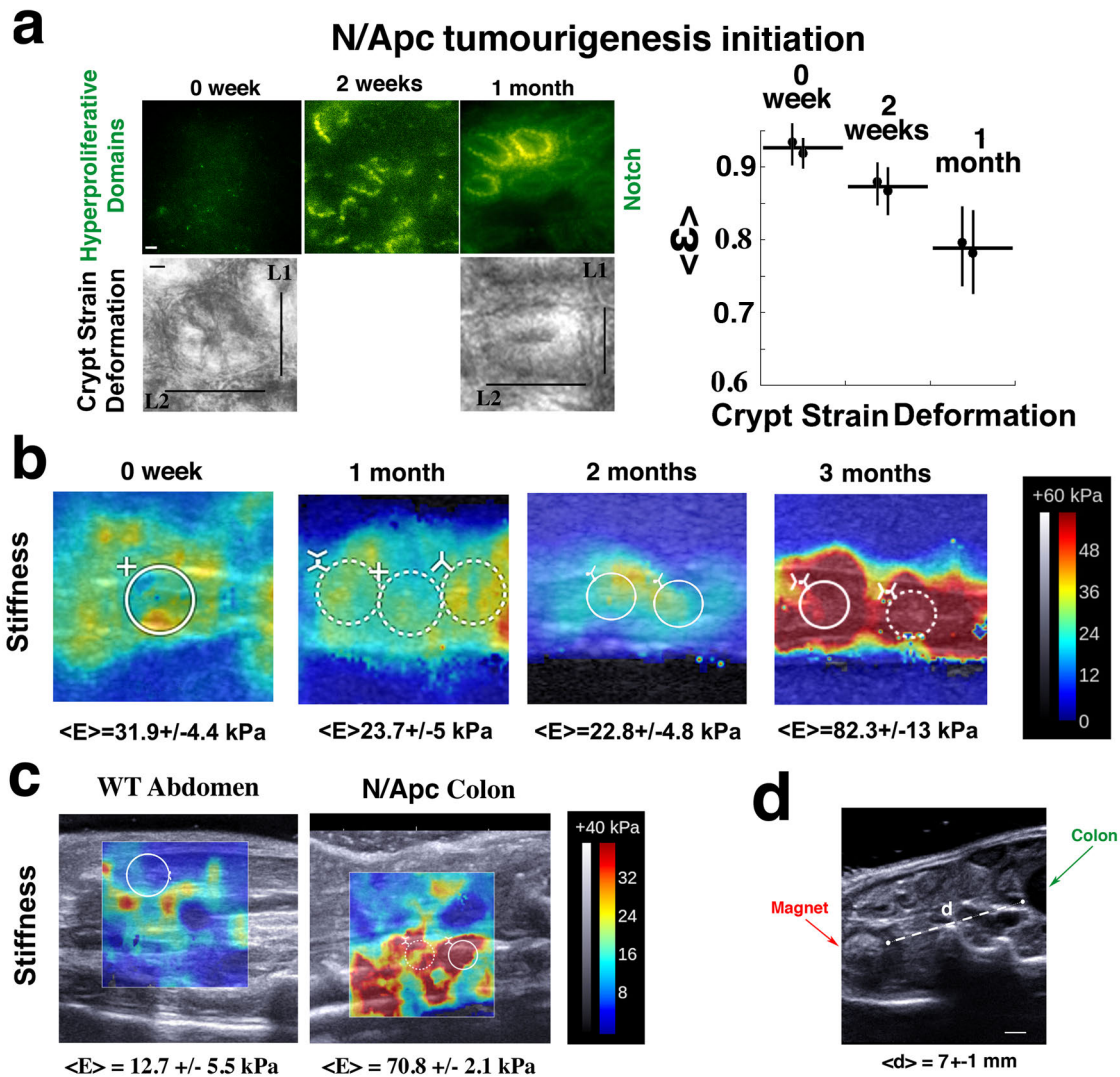
**Statistics.** All quantitative results were analysed using the non-parametric Mann-Whitney exact test and Student's *t*-test, both two-tailed, with standard deviation error bars being comparable within each group of data.

All experiments were reproduced at least twice in the laboratory. *n* is the number of measurements by experiment.

No randomization was performed. No blinding was performed. No statistical method or power analysis was used to predetermine sample size. The coordination of the distinct disciplinary aspects of the work (acoustic, magnetism, magnetic and tamoxifen injections, surgery, labelling, qPCR, etc.) between the ten groups involved was complex, and did not materially allow us to organize such procedures.

- Fodde, R. *et al.* A targeted chain-termination mutation in the mouse *Apc* gene results in multiple intestinal tumors. *Proc. Natl Acad. Sci. USA* **91**, 8969–8973 (1994).
- Massart, R. Preparation of aqueous magnetic liquids in alkaline and acidic media. *IEEE Trans. Magn.* **17**, 1247–1248 (1981).
- Szoka, F. Jr & Papahadjopoulos, D. Procedure for preparation of liposomes with large internal aqueous space and high capture by reverse-phase evaporation. *Proc. Natl Acad. Sci. USA* **75**, 4194–4198 (1978).
- Béalé, G. *et al.* Ultra magnetic liposomes for MR imaging, targeting, and hyperthermia. *Langmuir* **28**, 11834–11842 (2012).
- Martina, M. S. *et al.* Generation of superparamagnetic liposomes revealed as highly efficient MRI contrast agents for *in vivo* imaging. *J. Am. Chem. Soc.* **127**, 10676–10685 (2005).
- Martina, M. S. *et al.* Magnetic targeting of rhodamine-labeled superparamagnetic liposomes to solid tumors: *in vivo* tracking by fibered confocal fluorescence microscopy. *Mol. Imaging* **6**, 140–146 (2007).
- Plassat, V., Martina, M. S., Barratt, G., Ménager, C. & Lesieur, S. Sterically stabilized superparamagnetic liposomes for MR imaging and cancer therapy: pharmacokinetics and biodistribution. *Int. J. Pharm.* **344**, 118–127 (2007).
- Martina, M.-S., Wilhelm, C. & Lesieur, S. The effect of magnetic targeting on the uptake of magnetic-fluid-loaded liposomes by human prostatic adenocarcinoma cells. *Biomaterials* **29**, 4137–4145 (2008).
- Plassat, V. *et al.* Anti-estrogen-loaded superparamagnetic liposomes for intracellular magnetic targeting and treatment of breast cancer tumors. *Adv. Funct. Mater.* **21**, 83–92 (2011).
- Bain, J., McLauchlan, H., Elliott, M. & Cohen, P. The specificities of protein kinase inhibitors: an update. *Biochem. J.* **371**, 199–204 (2003).
- Blake, R. A. *et al.* SU6656, a selective src family kinase inhibitor, used to probe growth factor signaling. *Mol. Cell. Biol.* **20**, 9018–9027 (2000).
- Kim, D. W. *et al.* An orally administered multitarget tyrosine kinase inhibitor, SU1248, is a novel potent inhibitor of thyroid oncogenic RET/papillary thyroid cancer kinases. *J. Clin. Endocrinol. Metab.* **91**, 4070–4076 (2006).
- Langmuir, P. B. & Yver, A. Vandetanib for the treatment of thyroid cancer. *Clin. Pharmacol. Ther.* **91**, 71–80 (2012).
- Mologni, L., Redaelli, S., Morandi, A., Plaza-Menacho, I. & Gambacorti-Passerini, C. Ponatinib is a potent inhibitor of wild-type and drug-resistant gatekeeper mutant RET kinase. *Mol. Cell. Endocrinol.* **377**, 1–6 (2013).
- De Falco, V. *et al.* Ponatinib (AP24534) is a novel potent inhibitor of oncogenic RET mutants associated with thyroid cancer. *J. Clin. Endocrinol. Metab.* **98**, E811–E819 (2013).
- Gennisson, J. L. & Cloutier, G. Sol-gel transition in agar-gelatin mixtures studied with transient elastography. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **53**, 716–723 (2006).
- Chamming's, F., Latorre-Ossa, H., Lefrère-Belda, M. A., Fitoussi, V., Quibel, T., Assayag, F., Marangoni, E., Autret, G., Balvay, L., Pidial, L., Gennisson, J. L., Tanter, M., Cuenod, C. A., Clément, O. & Fournier, L. Shear wave elastography of tumor growth in a human breast cancer model with pathological correlation. *Eur. Radiol.* **23**, 2079–2086 (2013).
- Bercoff, J., Tanter, M. & Fink, M. Supersonic shear imaging: A new technique for soft tissue elasticity mapping. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **51**, 396–409 (2004).
- Sarvazyan, A. P., Rudenko, O. V., Swanson, S. D., Fowlkes, J. B. & Emelianov, S. Y. Shear wave elasticity imaging: A new ultrasonic technology of medical diagnostics. *Ultrasound Med. Biol.* **24**, 1419–1435 (1998).
- Swanson, C. A. *et al.* A comparison of diets of blacks and whites in three areas of the United States. *Nutr. Cancer* **20**, 153–165 (1993).

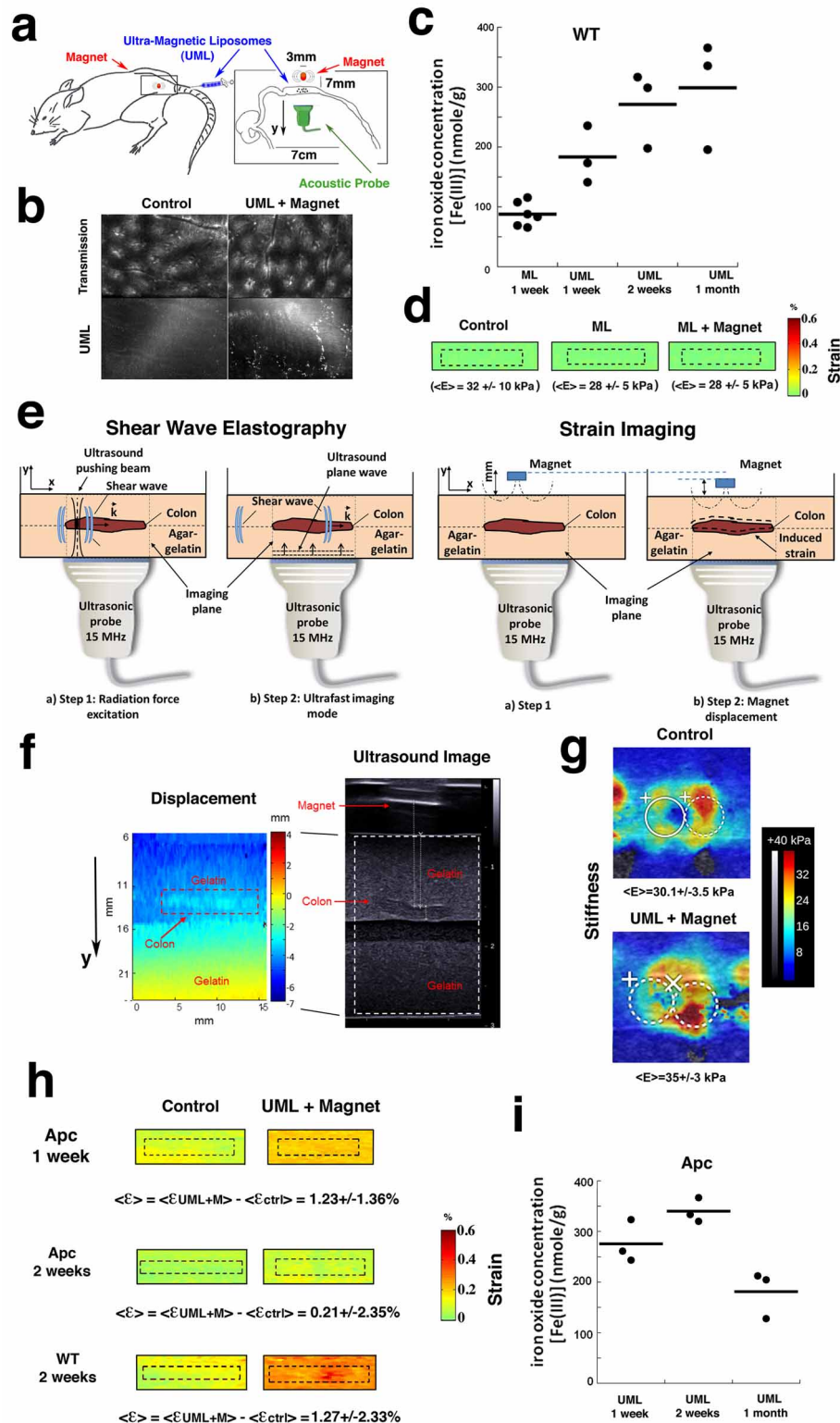
38. Goss, S. A., Johnston, R. L. & Dunn, F. Comprehensive compilation of empirical ultrasonic properties of mammalian tissues. *J. Acoust. Soc. Am.* **64**, 423–457 (1978).
39. Greenleaf, J. F., Fatemi, M. & Insana, M. Selected methods for imaging elastic properties of biological tissues. *Annu. Rev. Biomed. Eng.* **5**, 57–78 (2003).
40. Sarvazyan, A. *et al.* Biophysical bases of elasticity imaging. *Acoustical Imaging* **21**, 223–240 (1995).
41. Gennisson, J. L. *et al.* Viscoelastic and anisotropic mechanical properties of *in vivo* muscle tissue assessed by supersonic shear imaging. *Ultrasound Med. Biol.* **36**, 789–801 (2010).
42. Royer, D., Gennisson, J. L., Deffieux, T. & Tanter, M. On the elasticity of transverse isotropic soft tissues. *J. Acoust. Soc. Am.* **129**, 2757–2760 (2011).
43. Gennisson, J. L., Grenier, N., Combe, C. & Tanter, M. Supersonic shear wave elastography of *in vivo* pig kidney: influence of blood pressure, urinary pressure and tissue anisotropy. *Ultrasound Med. Biol.* **38**, 1559–1567 (2012).
44. Bercoff, J. *et al.* in *2008 IEEE Ultrasonics Symposium, Vols 1-4 and Appendix Ultrasonics Symposium* 321–324 (2008).
45. Madsen, E. L. *et al.* Anthropomorphic breast phantoms for testing elastography systems. *Ultrasound Med. Biol.* **32**, 857–874 (2006).
46. Ophir, J., Cespedes, I., Ponnekanti, H., Yazdi, Y. & Li, X. Elastography — a quantitative method for imaging the elasticity of biological tissues. *Ultrason. Imaging* **13**, 111–134 (1991).
47. Latorre-Ossa, H., Gennisson, J.-L., De Broesses, E. & Tanter, M. Quantitative imaging of nonlinear shear modulus by combining static elastography and shear wave elastography. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **59**, 833–839 (2012).
48. Righetti, R. *et al.* The feasibility of using poroelastographic techniques for distinguishing between normal and lymphedematous tissues *in vivo*. *Phys. Med. Biol.* **52**, 6525–6541 (2007).
49. Schmittgen, T. D. & Livak, K. J. Analyzing real-time PCR data by the comparative  $C_T$  method. *Nature Protocols* **3**, 1101–1108 (2008).
50. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acid Res.* **38**, e164 (2010).
51. Levy, M. *et al.* Long term *in vivo* biotransformation of iron oxide nanoparticles. *Biomaterials* **32**, 3988 (2011).



**Extended Data Figure 1 | Mechanical characterisation of pressure and stiffness in colonic tumours in Notch/Apc mice.** **a**, Crypt strain deformation analysis. Imaging of tumour initiation in wide-open colon samples. Time-dependent elliptical crypt shape change in Notch/Apc mice tissue explants analysed at 0, 2 weeks and 1 month after tamoxifen injection initiating mitotic Notch expression (Methods) (measured in  $n = 2$  mice per time point). Notch expression is revealed by GFP. Crypt strain deformation  $\epsilon = L1/((L1+L2)/2)$  ( $L1$  being the smallest radius, and  $L2$  the largest radius of the crypts elliptically deformed by the Notch dependent hyperproliferation tumour pressure induced), analysis using ImageJ.  $\epsilon_{0\text{week}} = 0.93 \pm 0.05$  (measured in  $n = 20$  crypts, two mice) for control sample at 0 weeks,  $\epsilon_{2\text{weeks}} = 0.874 \pm 0.06$  (measured in  $n = 31$  crypts, two mice,  $P = 0.029$ ) at 2 weeks, and  $\epsilon_{1\text{month}} = 0.79 \pm 0.11$  (measured in  $n = 20$  crypts, two mice,  $P < 0.0001$ ), at 1 month in colon explants. Statistic test is the non-parametric exact Mann–Whitney test.

Error bars are standard deviation. Scale bars, 5  $\mu\text{m}$ . **b**, Stiffness ( $E$ ) of distal colon explants embedded in agar–gelatin phantoms (Methods) measured by using shear wave elastography (see Extended Data Fig. 2e). Average values of  $E$  were obtained from at least  $n = 6$  measurements (in two mice for 0 weeks, 1 month and 2 months, three mice for 3 months). **c**, *In vivo* stiffness images of wild-type abdomen, and of Notch/Apc tumorous colon after 3 months of tumour development ( $E = 70.8 \pm 2.1$  kPa). Elasticity measurements were taken *in vivo* over anaesthetized mice (average of at least  $n = 6$  measurements, in two mice in each condition). **d**, Acoustic imaging distance of the magnet to the colon. The distance between the colon, observed in black after injection of echographic water gel in the lumen (Methods), and the round 3 mm diameter magnet, visualized as a result of acoustic refraction, was measured to be  $d = 7 \pm 1$  mm (measured in  $n = 4$  mice). The white bar is 1 mm.

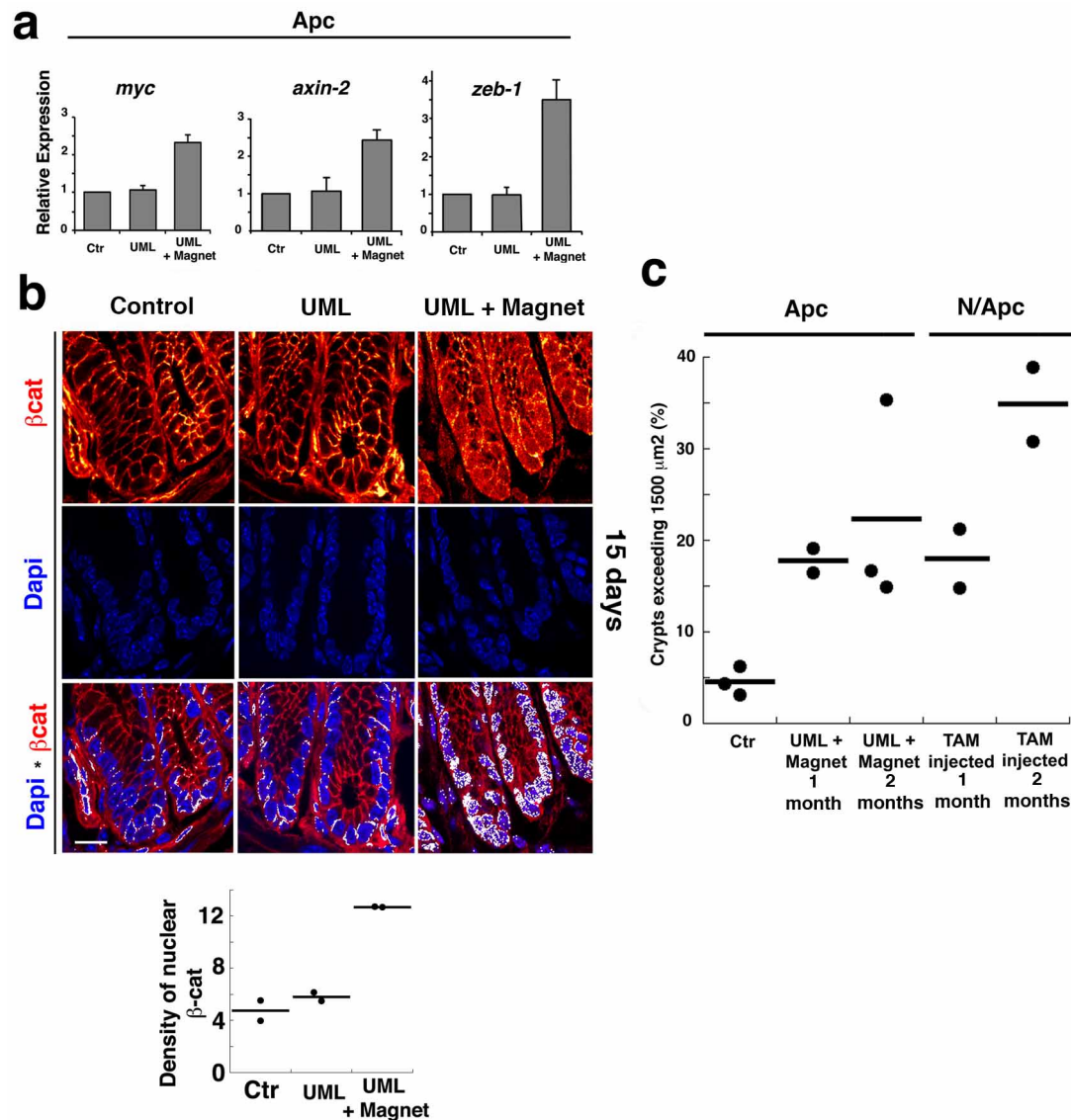




**Extended Data Figure 2 | Magnetic loading, iron oxide quantification by electronic spin resonance and strain compression calculation of magnetized wild-type and *Apc<sup>+/-1638N</sup>* colon distal samples.**

**a**, Magnetizing mouse colon. Subcutaneous insertion of the 3 mm diameter magnet (in red) on the back of the mice 7 mm in front of the distal colon, following UML injection in the lateral caudal vein of the mouse tail. Acoustic elastography probe in green. **b**, UML detection and localization in the colon. Rhodamine-labelled UML observed 1 week after injection in wild-type colon explants, in contrast to non-injected controls (measured in  $n = 10$  images, in six mice for each condition). **c**, Iron oxide quantification by electronic spin resonance in wild-type mice, injected with UML or magnetic liposomes loaded with an equivalent of 120 moles and 2.2 moles of Fe(III) per mole of lipids, respectively. Iron oxide concentration in the distal colon was measured by electronic spin resonance (5% precision) from lyophilized colon explants at 1 week (magnetic liposomes:  $79 \pm 12$  nmole  $g^{-1}$  (measured in  $n = 6$  mice); UML:  $183 \pm 48$  nmole  $g^{-1}$  (measured in  $n = 3$  mice), 2 weeks ( $271 \pm 64$  nmole  $g^{-1}$  (measured in  $n = 3$  mice)) and 1 month ( $299 \pm 157$  nmole  $g^{-1}$  (measured in  $n = 3$  mice)) after administration. Control sample was not injected with UML and was used to set up background of iron oxide concentration. A Mann–Whitney–Wilcoxon test concluded no significant difference in iron oxide concentration during 1 month ( $P \geq 0.2$ ) in the case of the mice injected with UML. This supports the maintaining of the number and local containment of the delivered magnetic particles into the distal colon over this period of time, in agreement with the very slow biotransformation of such nanomagnets recently revealed *in vivo*<sup>51</sup>. The iron oxide concentration at 1 week in the distal colon of wild-type mice injected with magnetic liposomes was found around twice lower compared to samples injected with UML while the residual content tends to vanish from 2 weeks (results not shown). **d**, Magnetized wild-type colon samples injected with

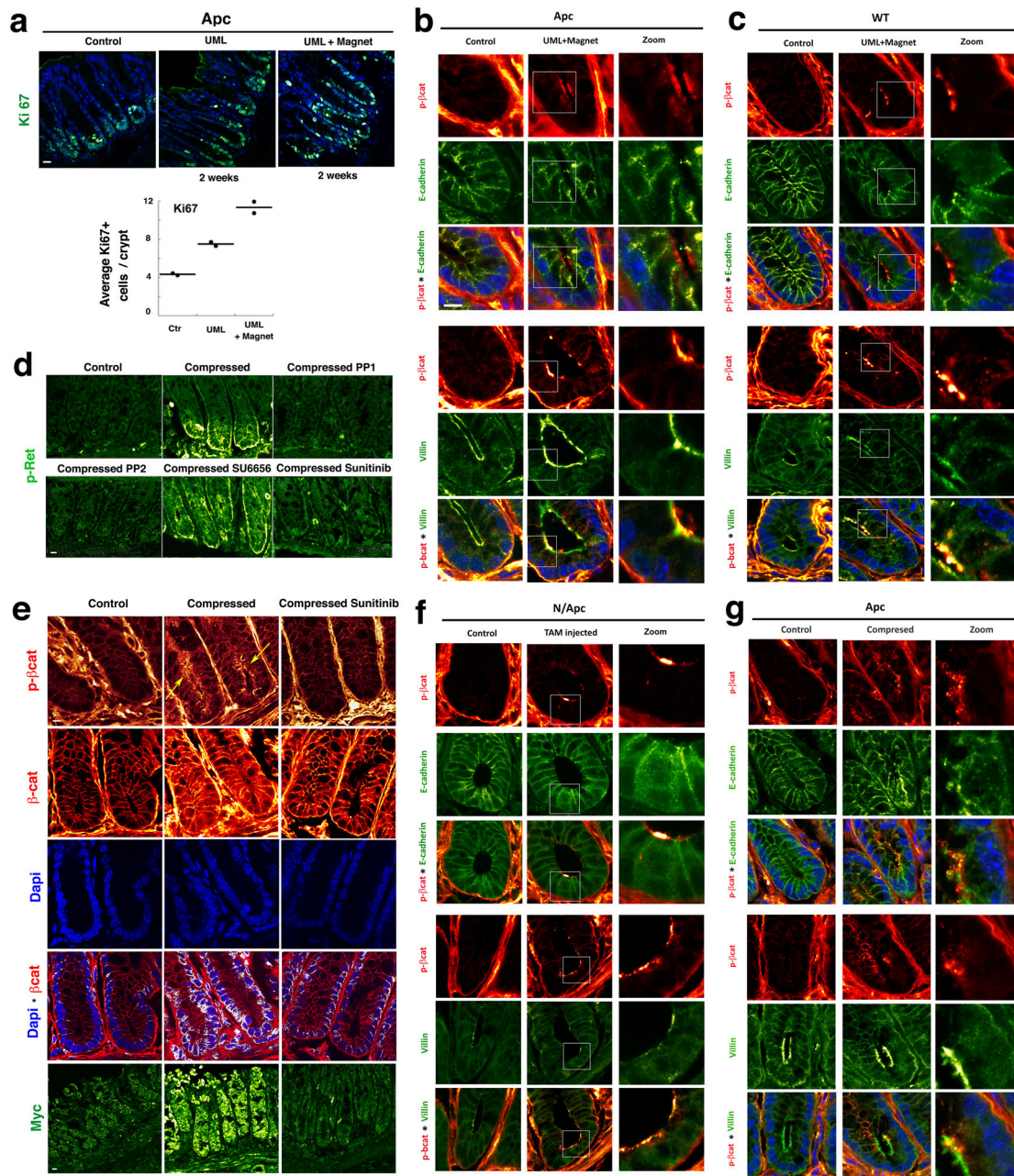
magnetic liposomes show no difference of local strain compression compared to non-injected samples or injected with magnetic liposomes without magnet (measured in  $n = 2$  for each condition). **e**, Schematic representation of the ultrasound measurement setup and biomechanical imaging techniques. Left, shear wave elastography. Right, strain imaging. **f**, Ultrasound imaging of magnetically induced deformation of the mouse colon. Representative B-mode acoustic image (right) of a magnetically loaded colon explant and the corresponding displacement map (left) after magnet moved from 10 mm to 7 mm towards the colon (Methods). The mean value of displacement within the colon and the rest of the phantom were  $35 \mu m$  and  $16.9 \mu m$ , respectively (measured in  $n = 12$  mice). **g**, Stiffness measurements of control (in 3 mice) and magnetized colon explants (in 6 mice) remained on the order of  $E = 30$ – $35$  kPa,  $n = 6$  measurements by mice. **h**, *Apc<sup>+/-1638N</sup>* mice show poor deformation at 1 week (mean elastic modulus control  $E = 24 \pm 2$  kPa, measured in  $n = 3$  mice; UML + magnet  $E = 23 \pm 2$  kPa, measured in  $n = 4$  mice) compared to wild type at 1 week (Fig. 1b) with similar iron oxide loading concentration of  $275 \pm 80$  nmole  $g^{-1}$  (see **i**,  $n = 3$ ) compared to the  $183 \pm 48$  nmole  $g^{-1}$  of wild type (see **c**). Deformation decreases at 2 weeks compared to 1 week in both the *Apc<sup>+/-1638N</sup>* (mean elastic modulus control  $E = 21 \pm 5$  kPa, measured in  $n = 3$  mice; UML + magnet  $E = 33 \pm 2$  kPa, measured in  $n = 2$  mice) and wild type (mean elastic modulus control  $E = 30 \pm 8$  kPa, measured in  $n = 3$  mice; UML + magnet  $E = 35 \pm 5$  kPa, measured in  $n = 3$  mice). **i**, The mean Fe(III) concentration in magnetized *Apc<sup>+/-1638N</sup>* colon, of  $265 \pm 80$  nmole  $g^{-1}$  on one month, remains on the order of magnitude of the mean concentration in wild-type colons, of  $250 \pm 50$  nmole  $g^{-1}$ , meaning that the magnetically induced stress remains on the same order of magnitude during 1 month in *Apc<sup>+/-1638N</sup>* compared to wild-type mice (measured in  $n = 3$  mice for each condition).



**Extended Data Figure 3 | Mechanical induction of  $\beta$ -catenin nuclear translocation, of target genes expression and ACF formation, though  $\beta$ -catenin Y654 phosphorylation in  $Apc^{+/1638N}$  mice colon.** **a**, Expression levels of *Myc*, *Axin2* and *Zeb1* observed by RT-qPCR (measured in  $n = 8$  mice for *Myc*,  $n = 9$  mice for *Axin2* and *Zeb1*). Non-significant  $P$  values between control and UML-injected conditions,  $P < 0.01$  between UML with magnet and UML-injected and control conditions (non-parametric exact Mann-Whitney test). Error bars are standard deviation. **b**, Nuclear  $\beta$ -catenin under UML with magnet conditions, compared to controls. White spots and purple represent a positive nuclear DAPI and  $\beta$ -catenin co-localization, with a preference for the peripheral privileged sites of transcriptionally active chromatin (measured in  $n = 2$  mice for each condition, ten images analysed with 4–6 crypts per image). Quantification of  $P$  values:  $P > 0.05$  for control

versus UML comparisons.  $P < 0.01$  for comparison between UML plus magnet versus UML without magnet, by Student's  $t$ -test. **c**, Number of large crypts ( $>1,500 \mu\text{m}^2$ ) in  $Apc^{+/1638N}$  mice under UML plus magnet conditions after 1 and 2 months (all crypts mean surface  $1,218.30 \pm 10.64$  and  $1,233.8 \pm 211.08 \mu\text{m}^2$ , 261 crypts, measured in  $n = 2$  mice and 195 crypts, measured in  $n = 3$  mice, respectively), compared to control without UML (mean surface  $927.6 \pm 124.72 \mu\text{m}^2$ , 545 crypts, measured in  $n = 3$  mice), and in Notch/Apc mice after 1 month and 2 months (mean surfaces of  $1,262.6 \pm 144.35$  and  $1,519.5 \pm 319.19 \mu\text{m}^2$ , 441 and 178 crypts, respectively, measured in  $n = 2$  mice for each condition), compared to non-tamoxifen-injected  $Apc^{+/1638N}$  control (mean surface  $927.6 \pm 124.72 \mu\text{m}^2$ , 545 crypts, measured in  $n = 3$  mice).  $P_{\text{Magnet1month}} < 0.01$ ,  $P_{\text{Magnet2months}} < 0.05$ ,  $P_{\text{Notch/Apc1month}} < 0.01$  and  $P_{\text{Notch/Apc2months}} < 0.001$  (analytic Student's  $t$ -test). Scale bars are  $10 \mu\text{m}$ .

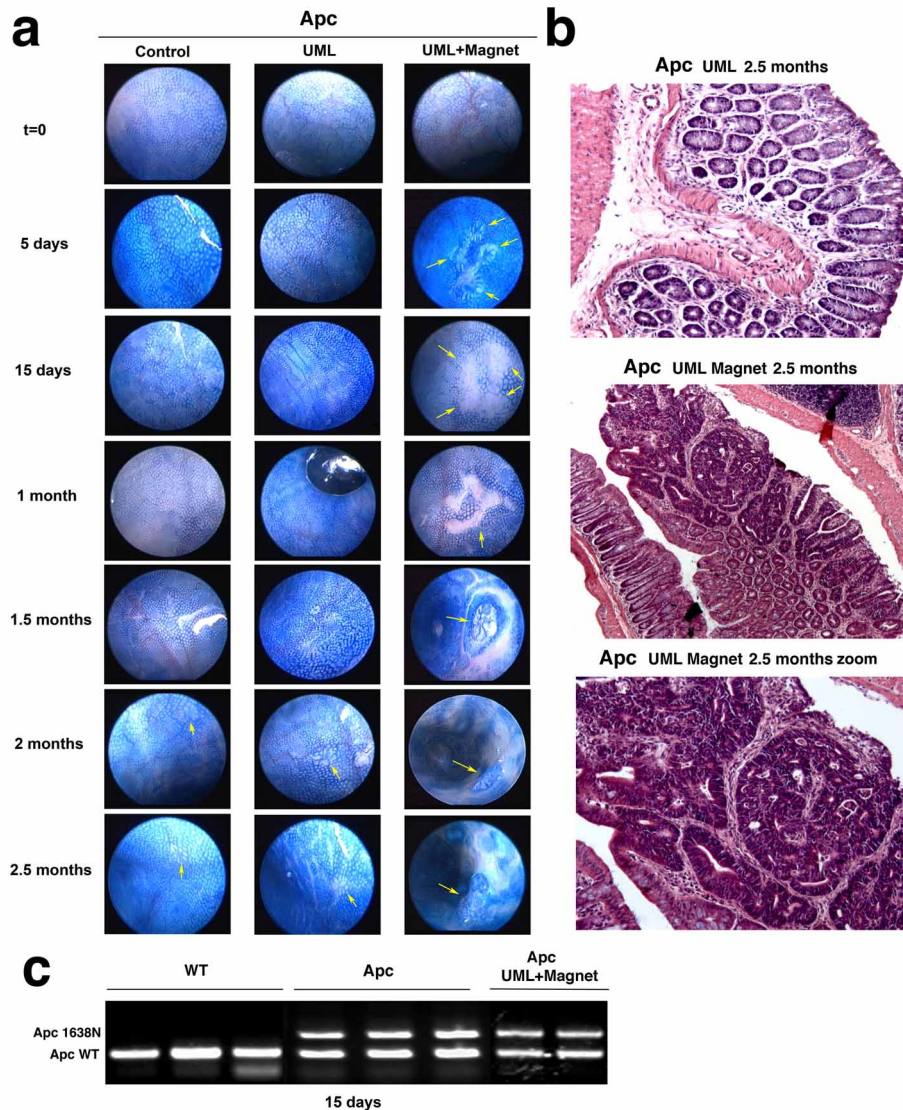




**Extended Data Figure 4 | Mechanically induced phospho-Tyr654- $\beta$ -catenin by tumour growth pressure do not co-localize with E-cadherin but with villin.**

**a**, Top, Ki67 expression under UML plus magnet conditions show increased number of positive cells per crypt, in contrast to controls (without UML, or without magnet with UML, measured in  $n = 2$  mice, 33 crypts, for each condition. Blue is DAPI. Bottom,  $P$  values are of  $P < 0.001$  between any individual conditions. Statistical test is the non-parametric exact Mann–Whitney test for all preceding quantitative experiments. **b**, No co-localization of pTyr654- $\beta$ -catenin with sub-apical junctional E-cadherin in response to tumour growth pressure and co-localization with villin, in  $Apc^{+/1638N}$  mice with magnetic pressure (measured in  $n = 2$  mice by condition). **c**, No co-localization of pTyr654- $\beta$ -catenin with sub-apical junctional E-cadherin in response to tumour growth pressure and co-localization with villin, in wild-type mice with magnetic pressure (measured in  $n = 2$  mice by condition). **d**, Immunofluorescence analysis of pY1062-Ret kinase in response to *ex vivo* mechanical strain. The phosphorylation of Ret Y1062 is observed in *ex vivo*  $Apc^{+/1638N}$  colon upon mechanical compression (Methods) ( $54.26 \pm 16\%$  positive crypts out of 236, measured in  $n = 7$  mice), initiating at 1 min compression and compared to control ( $0.6 \pm 1.3\%$  of positive crypts out of 99, measured in  $n = 5$  mice). Mechanical activation of pY1062-Ret is impaired by wide range PP1 and PP2 Src-family kinase inhibitors ( $1.55 \pm 0.25\%$  positive crypts out of 142 and  $6.03 \pm 3.5\%$  positive crypts out of 192, respectively, measured in  $n = 2$  mice for each condition). Sunitinib, a Ret inhibitor ( $9.85 \pm 3.8\%$  positive crypts out of 143, measured in  $n = 4$  mice), blocks mechanical activation of pY1062-Ret and SU6656, a Src-specific inhibitor ( $39.5 \pm 8.5\%$  of positive crypts out of 95, measured in  $n = 3$  mice), does not block mechanical

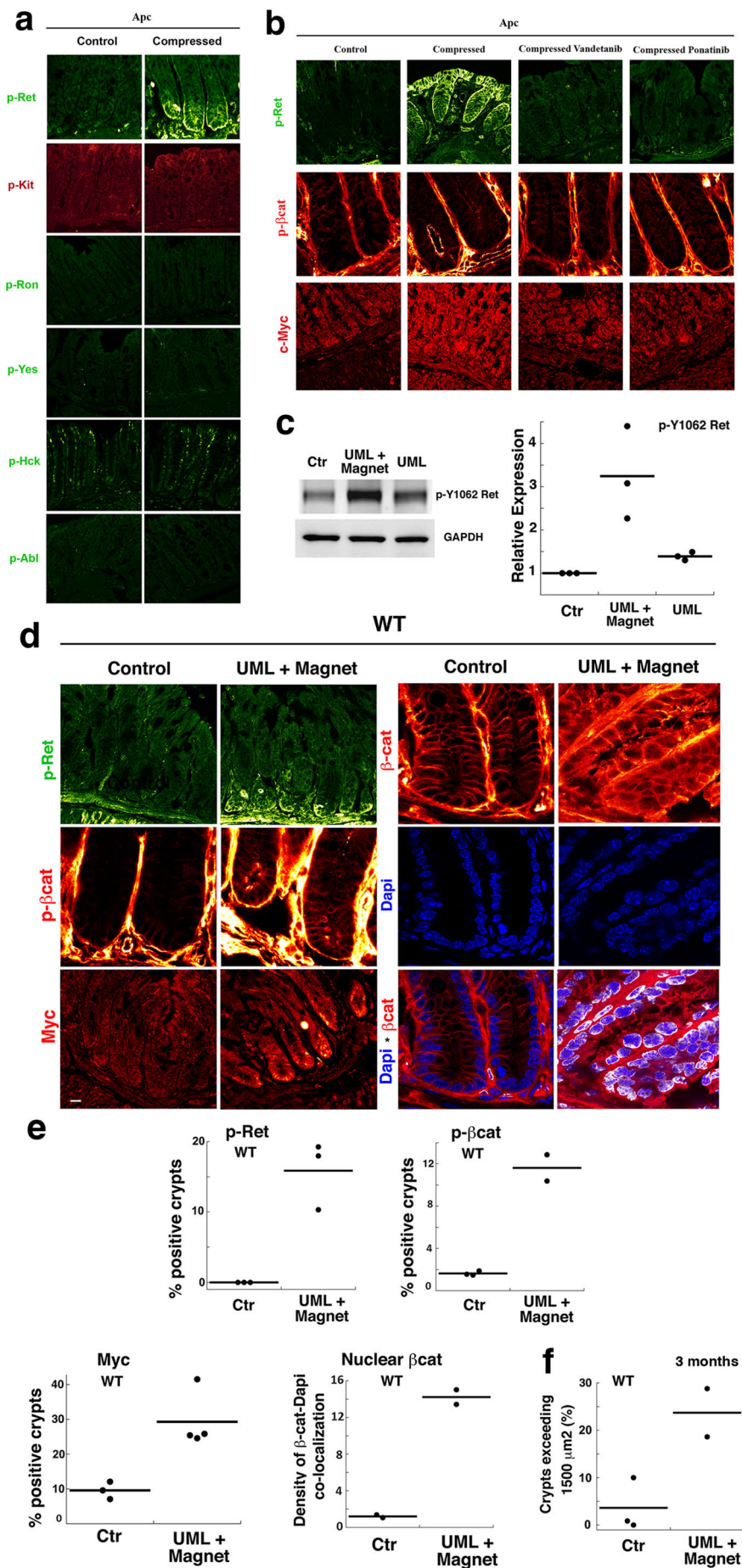
activation of pY1062-Ret. **e**, *Ex vivo* mechanical induction of pY654- $\beta$ -catenin after 20 min (measured in  $n = 2$  mice) (upper panel), enrichment of cytoplasmic and nuclear  $\beta$ -catenin after 20 min (measured in  $n = 3$  mice) (middle panel), and Myc expression after 4 h (measured in  $n = 6$  mice) (lower panel), impaired by Sunitinib inhibitor treatment (measured in  $n = 2$  mice by condition) in  $Apc^{+/1638N}$  mice.  $\beta$ -catenin control:  $13.97 \pm 2.7\%$  positive crypts out of 121, measured in  $n = 2$  mice; compressed:  $69.82 \pm 8.7\%$  positive crypts out of 119, measured in  $n = 2$  mice; compressed + sunitinib:  $24.13 \pm 1.2\%$  positive crypts out of 119, measured in  $n = 2$  mice. Myc control:  $10.96 \pm 0.21\%$  positive crypts out of 118, measured in  $n = 2$  mice; compressed:  $75.08 \pm 2.8\%$  positive crypts out of 148, measured in  $n = 6$  mice; compressed + sunitinib:  $23.36 \pm 1.1\%$  positive crypts out of 60, measured in  $n = 2$  mice. *Ex vivo* mechanical compression of  $Apc^{+/1638N}$  mice tissues caused an increase of  $\beta$ -catenin-positive nuclei ( $14.35 \pm 5$  a.u. (arbitrary units), measured in  $n = 2$  mice) by a factor of 3.6 compared to the control ( $3.97 \pm 1.9$  a.u., measured in  $n = 2$  mice), impaired in the presence of Sunitinib ( $4.9 \pm 1.9$  a.u., measured in  $n = 2$  mice). Merged images of DAPI and  $\beta$ -catenin were obtained with ImageJ co-localization analysis on ten images analysed with 4–6 crypts per image. White spots represent a positive co-localization. **f**, No co-localization of pTyr654- $\beta$ -catenin with sub-apical junctional E-cadherin and co-localization with villin in Notch/Apc mice (measured in  $n = 2$  mice). Here, the original colour for E-cadherin and villin was blue to avoid Notch–GFP green fluorescence, and was changed to green for the colour coherence of the complete figure. **g**, No co-localization of pTyr654- $\beta$ -catenin with sub-apical junctional E-cadherin and co-localization with villin, in  $Apc^{+/1638N}$  *ex vivo* compressed mice. Some apical E-cadherin is observed (measured in  $n = 2$  mice).



**Extended Data Figure 5 | Induction of ACF and adenoma formation in response to magnetic pressure mimicking tumour growth *in vivo* observed by colonoscopy in  $Apc^{+/1638N}$  mice.** **a**,  $Apc^{+/1638N}$  control: time 0, 0 aberrant crypt;  $t = 5$  days, 0 aberrant crypt;  $t = 15$  days,  $2 \pm 1$  ACF;  $t = 1$  month,  $4 \pm 1$  ACF;  $t = 1.5$  months,  $4 \pm 1$  ACF;  $t = 2$  months,  $5 \pm 2$  ACF;  $t = 2.5$  months,  $3 \pm 2$  ACF (measured in  $n = 3$  mice).  $Apc^{+/1638N}$  UML: time 0, 0 aberrant crypt;  $t = 5$  days, 0 aberrant crypt;  $t = 15$  days,  $2 \pm 1$  ACF;  $t = 1$  month,  $4 \pm 1$  ACF;  $t = 1.5$  months,  $4 \pm 3$  ACF;  $t = 2$  months,  $4 \pm 2$  ACF;  $t = 2.5$  months,  $3 \pm 1$  ACF (measured in  $n = 3$  mice). Mean size of ACF remained constant and around  $874.5 \pm 414 \mu m^2$  in both cases.  $Apc^{+/1638N}$  UML + magnet: time 0, 0 aberrant crypt;  $t = 5$  days,  $2 \pm 1$  ACF of size  $1,146.9 \pm 101 \mu m^2$ ;  $t = 15$  days,  $8 \pm 2$  ACF of size  $1,777.7 \pm 775 \mu m^2$  plus one small adenoma of  $4,229.38 \mu m^2$  size;  $t = 1$  month,  $8 \pm 2$  ACF of mean size  $3,483.9 \pm 665 \mu m^2$  plus one small adenoma;  $t = 1.5$  months,  $8 \pm 2$  ACF plus one bigger adenoma of  $6,379.92 \mu m^2$  size;  $t = 2$  months,  $8 \pm 2$  ACF plus one small adenoma of  $3,512.54 \mu m^2$  size plus one bigger adenoma;  $t = 2.5$  months,  $8 \pm 2$  ACF plus one small adenoma plus one

bigger adenoma of  $6,523.28 \mu m^2$  size (measured in  $n = 3$  mice). Time 0 corresponds to 4-month-old animals. **b**, Histologic characterization of the adenoma carcinoma induced by magnetic mechanical pressure mimicking tumour growth in the  $Apc^{+/1638N}$  mice using haematoxylin and eosin staining.  $Apc^{+/1638N}$  UML 2.5 months, no adenoma carcinoma is observed (measured in  $n = 3$  mice);  $Apc^{+/1638N}$  UML magnet 2.5 months and zoom, the crypt fusion and loss of apico-basal polarity in nucleus position show the carcinoma nature of the two adenoma observed (measured in  $n = 3$  mice). All mice were injected at 4 months. **c**, LOH (loss of heterozygosity) evaluation in colonic tissues of  $Apc^{+/1638N}$  mice under UML mechanical pressure by genotyping at 15 days. PCR analysis of DNA derived from  $Apc^{+/1638N}$  mice with oligonucleotides detecting the wild-type and mutant  $Apc^{1638N}$  allele. Wild-type mice show no  $Apc^{1638N}$  allele (measured in  $n = 3$  mice), whereas  $Apc^{+/1638N}$  heterozygous mice show a ratio  $Apc:Apc^{1638N}$  of  $1.1 \pm 0.1$  (measured in  $n = 3$  mice).  $Apc^{+/1638N}$  mice with UML and magnet show a ratio  $Apc:Apc^{1638N}$  of  $1.1 \pm 0.1$  (measured in  $n = 2$  mice).

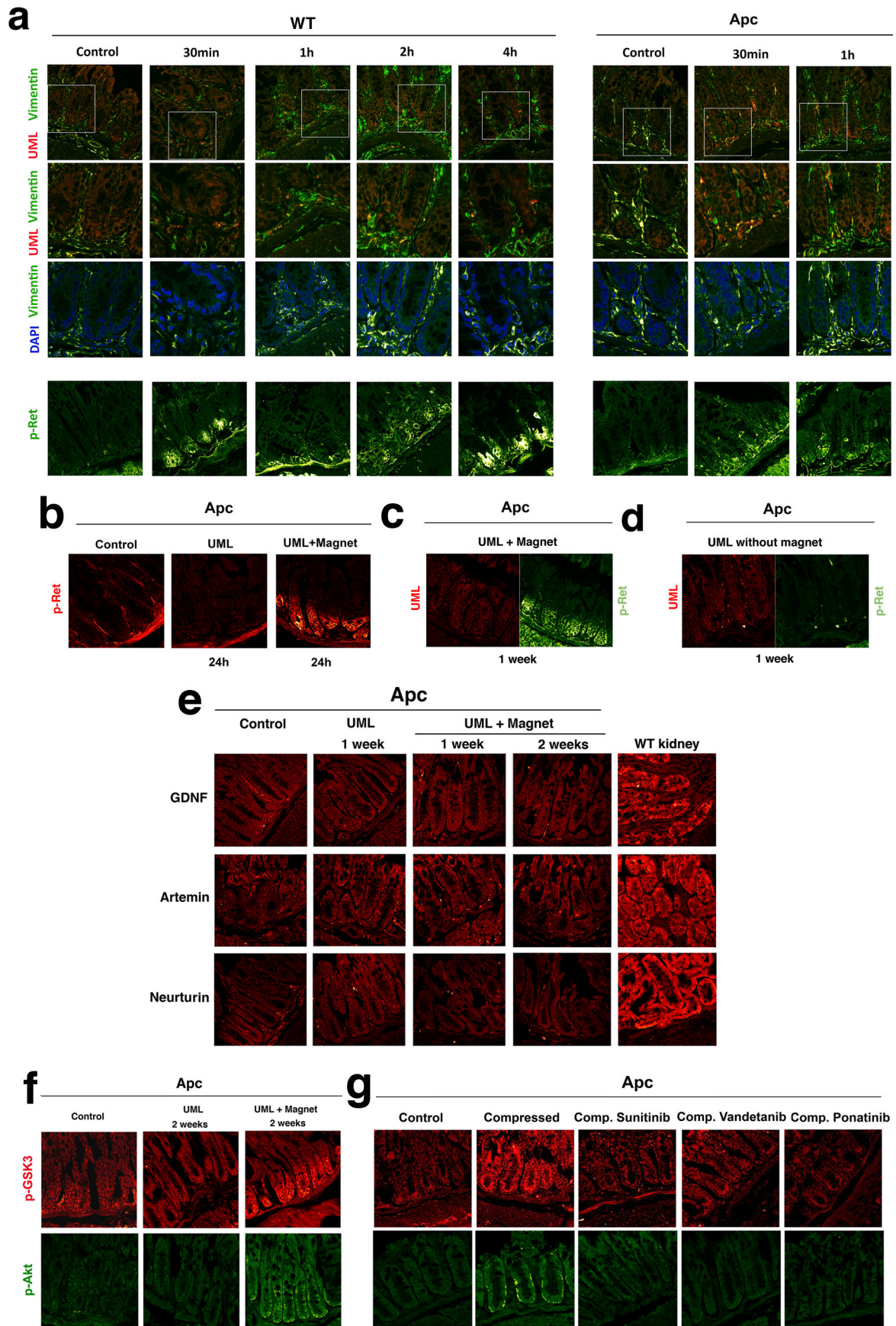




**Extended Data Figure 6 | Mechanical activation of Ret in both *Apc*<sup>+/1638N</sup> and wild-type mice.** **a**, Screening of the mechanical activation of the Src family kinases known to phosphorylate  $\beta$ -catenin Tyr654 and to be mechanosensitive in cell culture. Ret activation by Ret Tyr1062 phosphorylation under uni-axial compression (see Extended Data Fig. 4d legend for statistics). No phosphorylation of the Tyr568/570 site of activation of Kit under uni-axial compression. Control samples showed  $6.3 \pm 1.4\%$  positive crypts (258 total crypts observed, measured in  $n = 4$  mice) and compressed samples  $6.3 \pm 1.2\%$  positive crypts (399 total crypts observed, measured in  $n = 6$  mice). No phosphorylation of the Tyr1238/1239 site of activation of Ron under uni-axial compression. All negative crypts both in compressed samples (306 total crypts observed, measured in  $n = 4$  mice) and controls (347 total crypts observed, measured in  $n = 4$  mice). No phosphorylation of the Tyr537 site of activation of Yes. Control samples showed  $20.8 \pm 1.3\%$  positive crypts (307 total crypts observed, measured in  $n = 4$  mice) and compressed samples showed  $15.7 \pm 1.9\%$  positive crypts (312 total crypts observed, measured in  $n = 4$  mice). Positive crypts had only one positive stained cell. No significant phosphorylation increase of the Tyr209/211 site of activation of Hck under uni-axial compression (measured in  $n = 4$  mice for each condition).  $89.2 \pm 6.6\%$  positive crypts in compressed samples (268 total crypts observed, measured in  $n = 4$  mice) compared to control  $88.3 \pm 2.6\%$  positive crypts (258 total crypts observed, measured in  $n = 4$  mice). No phosphorylation of the Tyr412 site of activation of Abl under uni-axial compression. All negative crypts both in compressed samples (326 total crypts observed, measured in  $n = 4$  mice) and controls (332 total crypts observed, measured in  $n = 4$  mice). **b**, Inhibition of pTyr1062-Ret, pTyr654- $\beta$ -catenin and Myc mechanical induction by the two additional inhibitors of Ret, vandetanib and ponatinib. Top, Tyr1062 phosphorylation of Ret under uni-axial compression ( $54.16 \pm 15\%$  positive crypts of 236,  $n = 7$  mice), initiating at 1-min compression, compared to control ( $0.6 \pm 0.8\%$  positive crypts of 99, measured in  $n = 5$  mice) is inhibited in the presence of vandetanib ( $7.91 \pm 1.2\%$  positive crypts of 128 total crypts observed (measured in  $n = 2$  mice)) and ponatinib ( $5.17 \pm 1.3\%$  positive crypts out of 193 total crypts observed (measured in  $n = 2$  mice)). Middle, initiation of the  $\beta$ -catenin oncogenic pathway by phosphorylation of  $\beta$ -catenin on Tyr654

under uni-axial compression ( $69.82 \pm 8.7\%$  positive crypts of 119, measured in  $n = 2$  mice) compared to control ( $13.97 \pm 2.7\%$  positive crypts of 121, measured in  $n = 2$  mice) is inhibited in the presence of vandetanib ( $5.89 \pm 4.2\%$  positive crypts of 102 total crypts observed, measured in  $n = 2$  mice) and ponatinib ( $9.13 \pm 3.3\%$  positive crypts of 120 total crypts observed, measured in  $n = 2$  mice). Bottom, mechanical induction of Myc under uni-axial compression ( $50.5 \pm 0.2\%$  positive crypts of 146, measured in  $n = 2$  mice) compared to control ( $11 \pm 0.2\%$  positive crypts of 118, measured in  $n = 2$  mice), is altered in the presence of vandetanib ( $36.7 \pm 2.2\%$  positive crypts of 108 total crypts observed, measured in  $n = 2$  mice) and ponatinib ( $20.2 \pm 3.5\%$  positive crypts of 182 total crypts observed, measured in  $n = 2$  mice). The concentrations used were  $10 \mu\text{M}$  for vandetanib, and  $100 \text{ nM}$  for Ponatinib. **c**, pY1062-Ret induction quantified by western blot. Control (measured in  $n = 3$  mice), UML + magnet (measured in  $n = 3$  mice) and UML alone (measured in  $n = 3$  mice) conditions.  $P > 0.05$  between control and UML,  $P < 0.01$  between UML and UML + magnet, Student *t*-test. **d**, Immunofluorescence analysis of wild-type and UML-injected colon in the presence of the magnet of pY1062-Ret (at 2 weeks; control: measured in  $n = 3$  mice, 300 crypts minimum; UML + magnet: measured in  $n = 3$  mice, 469 crypts), of pY654- $\beta$ -catenin apically (at 1 month; control: measured in  $n = 3$  mice, 798 crypts; UML + magnet: measured in  $n = 2$  mice, 495 crypts), of cytoplasmic and nuclear enrichment of  $\beta$ -catenin (at 2 months; control: 14 images analysed with 4–6 crypts per image; UML + magnet: ten images analysed with 4–6 crypts per image; measured in  $n = 2$  mice for each condition), and of the expression of Myc (at 2 months; control: measured in  $n = 3$  mice, 245 crypts; UML + magnet: measured in  $n = 4$  mice, 453 crypts). White spots and purple represent positive co-localization, with a preference for peripheral sites. **e**, Quantification of **d**. *P* values: pRet ( $P < 0.001$ ), p $\beta$ -catenin ( $P < 0.001$ ),  $\beta$ -cat ( $P < 0.001$ ) and Myc ( $P < 0.001$ ), by Student's *t*-test. **f**, Percentage of crypts exceeding  $1,500 \mu\text{m}^2$  of mean surface area 3 months after UML injection mimicking tumour growth pressure (mean surface  $1,311.29 \pm 82 \mu\text{m}^2$ , measured in  $n = 2$  mice, 285 crypts) compared to the control without UML (mean surface  $914.3 \pm 189.09 \mu\text{m}^2$ , measured in  $n = 3$  mice, 432 crypts).  $P_{\text{Magnet3month}} < 0.05$  (analytic Student's *t*-test). Scale bars are  $10 \mu\text{m}$ .







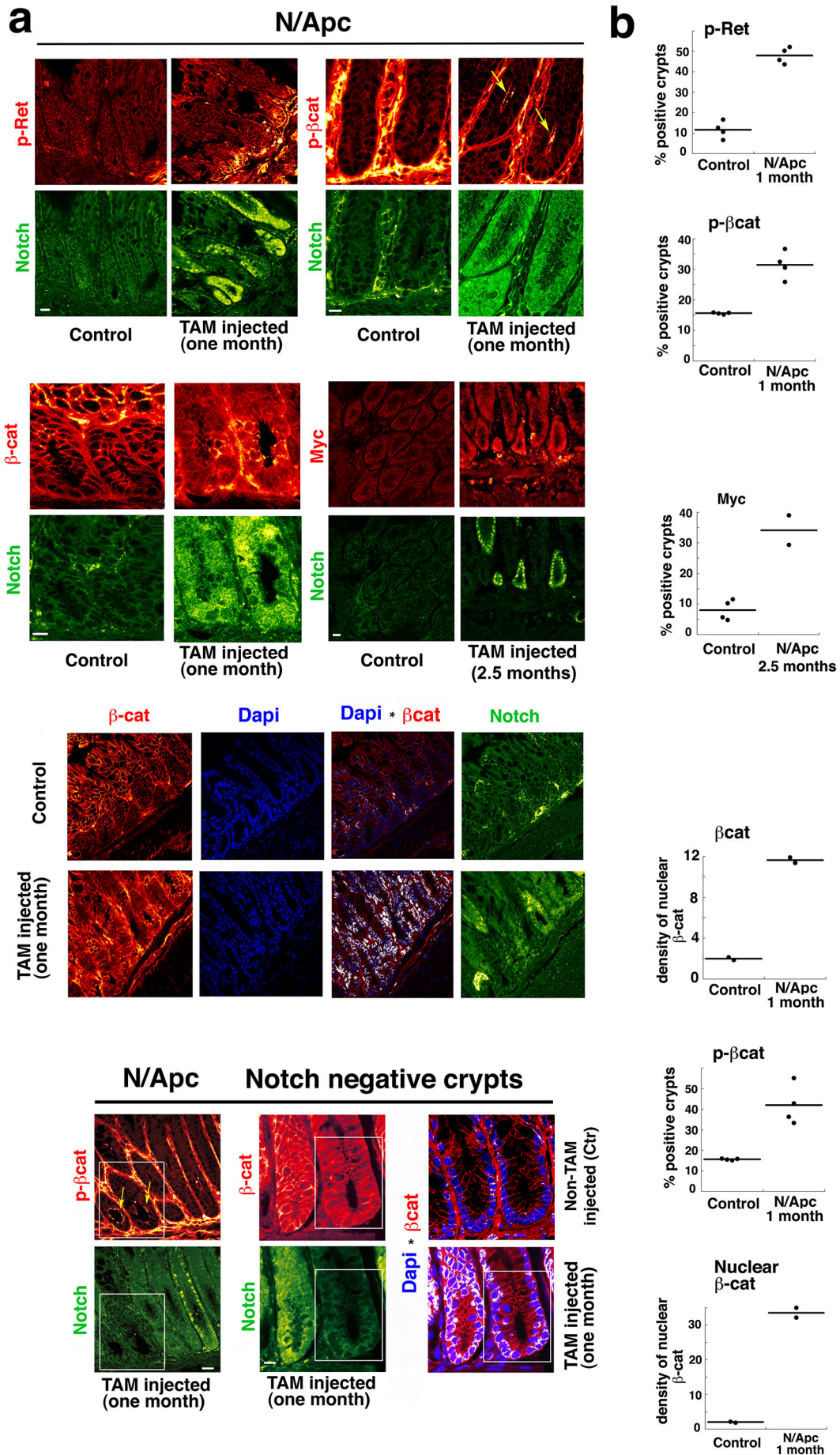
**Extended Data Figure 7 | Ret activation is rapidly and non-cell****autonomously induced by the mechanical strains developed by magnetic**

**pressure.** **a,** Rapid mechanical activation of pTyr1062-Ret initiating 30 min after UML injection in the presence of magnet in wild-type and in *Apc*<sup>+/-1638N</sup> mice. Wild-type control: all negative crypts. Wild type + UML + magnet 30 min: 20.2 ± 0.56% positive crypts (measured in *n* = 2 mice, 372 crypts), 1 h: 44.56 ± 22.81% positive crypts (measured in *n* = 2 mice, 377 crypts), 2 h: 61.4 ± 24.12% positive crypts (measured in *n* = 2 mice, 458 crypts), 4 h: 67.5 ± 10.23% positive crypts (measured in *n* = 2 mice, 510 crypts). *Apc*<sup>+/-1638N</sup> control: 1.22 ± 0.48% positive crypts (measured in *n* = 4 mice, 556 crypts). *Apc*<sup>+/-1638N</sup> + UML + magnet 30 min: 20.18 ± 3.73% positive crypts (measured in *n* = 2 mice, 417 crypts), 1 h: 28.77 ± 7.87% positive crypts (measured in *n* = 2 mice, 382 crypts). **b,** Maintenance of mechanical activation of Tyr1062-Ret 24 h after UML injection in the presence of magnet in *Apc*<sup>+/-1638N</sup> mice. *Apc*<sup>+/-1638N</sup> control: 1.22 ± 0.5% pTyr1062-Ret-positive crypts out of 556 total crypts (measured in *n* = 4 mice). *Apc*<sup>+/-1638N</sup> + UML (without magnet): 3.88 ± 0.6% of pTyr1062-Ret-positive crypts out of 568 total crypts (measured in *n* = 2 mice). *Apc*<sup>+/-1638N</sup> + UML + magnet: 14.15 ± 2.5% of pTyr1062-Ret-positive crypts out of 499 total crypts (measured in *n* = 2 mice). **c,** Mechanical activation of pTyr1062-Ret kinase in non-UML-loaded local domains, after UML injection and magnet implantation. A strong phosphorylation of Ret Tyr1062 (green) was observed not only in the domains where UML were accumulated (not shown) but also in UML-absent domains (negative signal for fluorescent rhodamine, left) in *Apc*<sup>+/-1638N</sup> colons (measured in *n* = 4 mice for each condition). **d,** No mechanical activation of pTyr1062-Ret kinase after UML injection in the absence of magnet in *Apc*-deficient colons. In some domains where an accumulation of UML in the conjunctive tissue could be observed (left) no Ret phosphorylation was observed (right) (measured in *n* = 3 mice for each condition). **e,** Expression of GDNF, artemin, and neurturin do not change in magnetized *Apc*<sup>+/-1638N</sup> colon explants. We

observed no effect of magnetization of the tissue in the expression of any ligand of Ret as compared to the control not injected with UML, to the non-magnetized colon tissue injected with UML and to the wild-type kidney positive control expressing the three ligands of Ret (minimum of 300 crypts analysed, measured in *n* = 2 mice for each condition).

**f,** Mechanical inactivation of GSK-3β through Ser9 phosphorylation and mechanical activation of the upstream Akt through Ser473 phosphorylation in magnetized *Apc*-deficient colon explants. Increased phosphorylation of GSK-3β Ser9 was observed in the magnetized tissue at 2 weeks (33.8 ± 0.4% of 307 crypts), compared to the non-magnetized control (12.7 ± 5.15% of 560 crypts) and to the UML-injected colon sample without magnet (7.24 ± 2% of 138 crypts) (measured in *n* = 2 mice for each condition). Increased phosphorylation of Akt Ser473 was observed in the magnetized tissue at 2 weeks (40.6 ± 4.5% of 378 crypts, measured in *n* = 3 mice), compared to the non-magnetized control (3.2 ± 1.1% of 596 crypts, measured in *n* = 3 mice) and to the UML-injected colon sample without magnet (2.7 ± 0.2% of 225 crypts, measured in *n* = 2 mice).

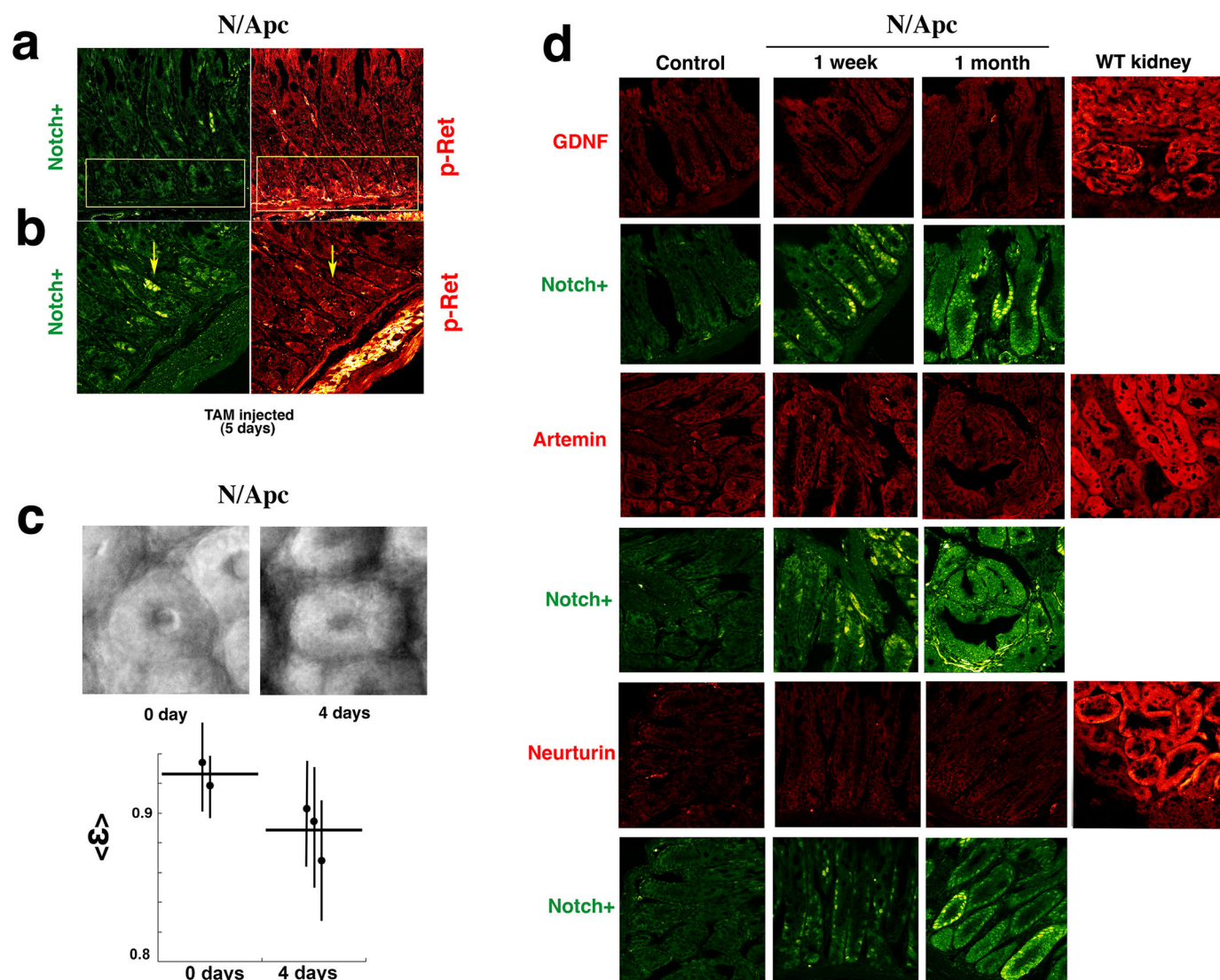
**g,** Ret-dependent mechanical inactivation of GSK-3β and activation of the upstream Akt after *ex vivo* global compression of *Apc*-deficient colon explants. GSK-3β: control, 12.5 ± 5.2% positive crypts out of 320 (measured in *n* = 2 mice); compressed, 43.26 ± 11% positive crypts out of 784 (measured in *n* = 2 mice); compressed + sunitinib, 19.7 ± 12% positive crypts out of 216 (measured in *n* = 2 mice); compressed + vandetanib, 14.4 ± 0.1% positive crypts out of 114 (measured in *n* = 2 mice); compressed + ponatinib, 15.45 ± 5.6% positive crypts out of 230 (measured in *n* = 2 mice for each condition). Akt: control, 5.3 ± 4.3% positive crypts out of 519; compressed, 46.6 ± 3.5% positive crypts out of 554; compressed + sunitinib, 14.26 ± 8.3% positive crypts out of 560; compressed + vandetanib, 5.37 ± 0.5% positive crypts out of 571; compressed + ponatinib, 7.52 ± 0.8% positive crypts out of 443 (measured in *n* = 2 mice for each condition).



**Extended Data Figure 8 | Notch/Apc colon shows an activation of the Ret/ $\beta$ -catenin/Myc mechanotransductive signalling pathway one month after tumour growth initiation.** **a**, Top, Ret Y1062 phosphorylation (measured in  $n = 4$  mice, 317 crypts), apical  $\beta$ -catenin Y654 phosphorylation (measured in  $n = 4$  mice, 223 crypts), cytoplasmic and nuclear  $\beta$ -catenin enrichment (measured in  $n = 2$  mice, 16 images analysed with 4–6 crypts per image), and Myc expression activation (at 2.5 months, measured in  $n = 2$  mice, 194 crypts), in Notch/Apc tumorous tissue compared to non-tamoxifen-injected Notch/Apc mice controls (measured in  $n = 4$  mice, 264 crypts; measured in  $n = 4$  mice, 198 crypts; measured in  $n = 2$  mice, 18 images analysed with 4–6 crypts per image; measured in  $n = 2$  mice, 315 crypts, respectively). GFP fluorescence (green) reveals a nuclear overexpression of Notch in the tumorous crypts. Note that at 1 month after tamoxifen injection, GFP expression is often diffuse and found in the nuclei and cytoplasm. Immunofluorescence staining and ImageJ co-localization analysis revealed an enrichment of nuclear  $\beta$ -catenin (white and purple spots represent a positive co-localization between  $\beta$ -catenin (red) and DAPI (blue)) by a factor of 5.8 in Notch/Apc colon samples 1 month after tamoxifen injection ( $11.64 \pm 2.5$  a.u., measured in  $n = 2$  mice), compared to the control ( $1.99 \pm 0.8$  a.u., measured in  $n = 2$  mice). Bottom, phosphorylation of

$\beta$ -catenin Y654 in Notch-negative crypts in tamoxifen-injected crypts (measured in  $n = 4$  mice, 68 crypts, yellow arrows) compared to non-tamoxifen-injected control conditions (measured in  $n = 4$  mice, 198 crypts). Cytoplasmic enrichment and nuclear translocation of  $\beta$ -catenin in Notch-negative crypts (measured in  $n = 4$  mice). Nuclear translocation of  $\beta$ -catenin is assessed by co-localization with DAPI (in white) in Notch-negative crypts of Notch/Apc tissues after 1 month of tamoxifen injection compared to non-tamoxifen-injected control conditions. White spots and purple represent a positive nuclear DAPI and  $\beta$ -catenin co-localization, with a preference for the peripheral privileged sites of transcriptionally active chromatin. Immunofluorescence staining and ImageJ co-localization analysis revealed an enrichment of nuclear  $\beta$ -catenin (white and purple spots represent a positive co-localization between  $\beta$ -catenin (red) and DAPI (blue)) by a factor of 15 in Notch-negative crypts in colon samples 1 month after tamoxifen injection ( $33.6 \pm 8.7$  a.u., measured in  $n = 2$  mice), compared to the control ( $1.99 \pm 0.8$  a.u., measured in  $n = 2$  mice). Ten images with 4–6 crypts per image were analysed in each condition. Scale bars are 10  $\mu$ m. **b**, Quantification of **a**.  $P < 0.001$  in all cases, Student's  $t$ -test.





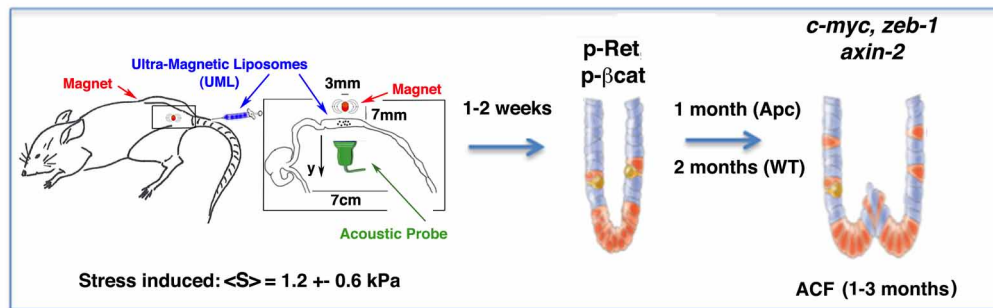
**Extended Data Figure 9 | Ret activation is non-cell autonomously induced by the mechanical strains developed by tumour pressure in Notch/Apc mice.**

**a**, Early mechanical activation of pTyr1062-Ret kinase in crypts totally surrounded by a Notch-negative domain in tumorous Notch/Apc colon explants. Notch/Apc mice were injected with tamoxifen for four consecutive days (instead of 5 days, see Methods) to induce tumour growth initiation, and colon explants were analysed on the fifth day by immunofluorescence. Ret phosphorylation was activated in  $12.7 \pm 2.3\%$  of Notch-negative crypts (22 pTyr1062-Ret-positive crypts of a total of 177 GFP-negative crypts) completely surrounded by Notch-negative crypts (GFP-negative), measured in  $n = 4$  mice. **b**, No phosphorylation of Ret Tyr1062 in Notch-overexpressing cells in early tumorous Notch/Apc colon explants.  $34.9 \pm 8.8\%$  of GFP-positive crypts (yellow arrows) showed no expression of pTyr1062-Ret (39 crypts out of a total of 113 crypts, measured in  $n = 4$  mice). **c**, Strain deformation of Notch/Apc 4

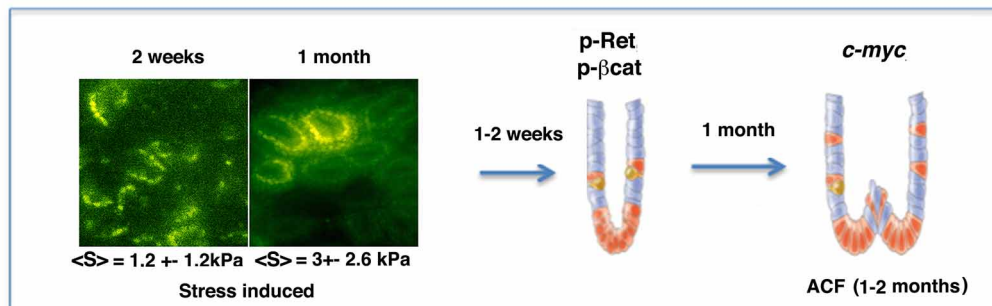
days after tamoxifen injection.  $\epsilon_{4d} = 0.89 \pm 0.08$  (measured in  $n = 169$  crypts, in three mice),  $\epsilon_{0d} = 0.93 \pm 0.05$  (measured in  $n = 20$  crypts in two mice),  $P = 0.004$  (Mann–Whitney exact test), leading to a mean tumour stress of  $S_{4d} = 0.9 \pm 0.1$  kPa (following Hooke's law  $S = E \times \epsilon$  with  $E_{\text{Notch/Apc}} = 22.8 \pm 4.8$  kPa (see Extended Data Fig. 1b)). Error bars are s.d. **d**, Expression of GDNF, artemin and neurturin Ret ligands does not change in tumour-initiated Notch/Apc colon. We observed no effect of tumour growth in the expression of any of the ligands (red) as compared to the control Notch/Apc not injected with tamoxifen, and to the wild-type kidney positive controls in which the three ligands of Ret are expressed. GFP fluorescence reveals positive expression of Notch and labels the tumorous crypts (minimum 300 crypts analysed, measured in  $n = 2$  mice for each condition). At 1 month after tamoxifen (tam.) injection, GFP expression is often diffuse and found in the nuclei and cytoplasm.

**a**

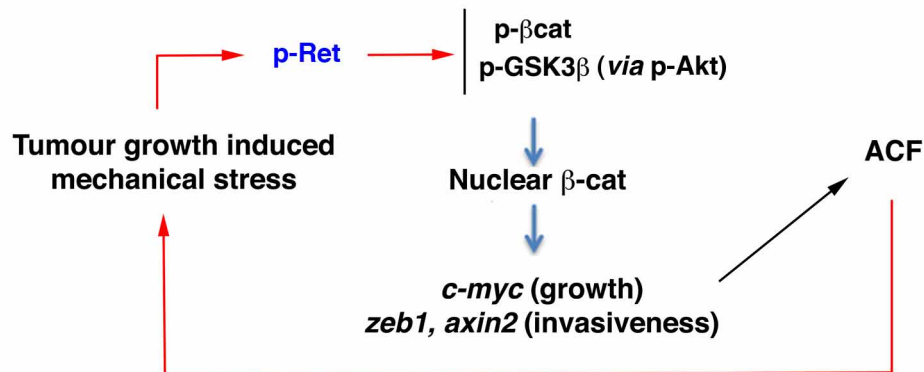
### In vivo magnetic induction of a mechanical tumour stress in heterozygous *APC* and WT is tumorigenic

**b**

### Notch hyperproliferative domains mechanical tumour stress induced in heterozygous *APC* domains is tumorigenic

**c**

### Mechano-genetic model of tumour growth instability



**Extended Data Figure 10 | Summary of the results.** a, Magnetically induced tumour growth pressure. b, Notch-induced tumour growth pressure. c, Ret/β-

catenin promote tumorigenic target gene expression and ACF formation and are mechanically induced by tumour growth pressure.

# MYC regulates the core pre-mRNA splicing machinery as an essential step in lymphomagenesis

Cheryl M. Koh<sup>1\*</sup>, Marco Bezzi<sup>1,2\*</sup>, Diana H. P. Low<sup>1</sup>, Wei Xia Ang<sup>1</sup>, Shun Xie Teo<sup>1</sup>, Florence P. H. Gay<sup>1</sup>, Muthafar Al-Haddawi<sup>1</sup>, Soo Yong Tan<sup>1</sup>, Motomi Osato<sup>3</sup>, Arianna Sabò<sup>4</sup>, Bruno Amati<sup>4,5</sup>, Keng Boon Wee<sup>6,7</sup> & Ernesto Guccione<sup>1,2,3</sup>

**Deregulated expression of the MYC transcription factor occurs in most human cancers and correlates with high proliferation, reprogrammed cellular metabolism and poor prognosis<sup>1</sup>. Overexpressed MYC binds to virtually all active promoters within a cell, although with different binding affinities<sup>2–4</sup>, and modulates the expression of distinct subsets of genes<sup>1,2,4,5</sup>. However, the critical effectors of MYC in tumorigenesis remain largely unknown. Here we show that during lymphomagenesis in Eμ-myc transgenic mice, MYC directly upregulates the transcription of the core small nuclear ribonucleoprotein particle assembly genes, including *Prmt5*, an arginine methyltransferase that methylates Sm proteins<sup>6,7</sup>. This coordinated regulatory effect is critical for the core biogenesis of small nuclear ribonucleoprotein particles, effective pre-messenger-RNA splicing, cell survival and proliferation. Our results demonstrate that MYC maintains the splicing fidelity of exons with a weak 5' donor site. Additionally, we identify pre-messenger-RNAs that are particularly sensitive to the perturbation of the MYC–PRMT5 axis, resulting in either intron retention (for example, *Dvl1*) or exon skipping (for example, *Atr*, *Ep400*). Using antisense oligonucleotides, we demonstrate the contribution of these splicing defects to the anti-proliferative/apoptotic phenotype observed in PRMT5-depleted Eμ-myc B cells. We conclude that, in addition to its well-documented oncogenic functions in transcription<sup>2–5</sup> and translation<sup>8</sup>, MYC also safeguards proper pre-messenger-RNA splicing as an essential step in lymphomagenesis.**

We recently provided an overview of the transcriptional networks perturbed by MYC during lymphomagenesis<sup>4</sup>: to this aim, we established MYC binding and gene expression profiles in B cells from non-transgenic control mice (C), as well as pre-tumoural B cells (P) and lymphomas (tumour, 'T') in Eμ-myc animals<sup>4</sup>. Among the gene sets upregulated by MYC, we identified components of the spliceosome (Fig. 1a). Recent genome-wide association studies have uncovered a high rate of mutations in splicing regulators<sup>9–11</sup>, pinpointing their potential involvement as driver oncogenes; consequently, there is growing interest in drugging the spliceosome machinery for anti-cancer therapy<sup>12</sup>. Among the significantly modulated genes involved in splicing, MYC promoted transcription of the core small nuclear ribonucleoprotein particle (snRNP) assembly genes, of which PRMT5 is the key enzymatic component<sup>7,13</sup> (Fig. 1b and Extended Data Fig. 1a): these genes showed progressive transcriptional increases in the pre-tumoural and tumour stages, as assessed by RNA sequencing (RNA-seq), validated by quantitative PCR (Fig. 1b, c and Extended Data Fig. 1b, c) and verified at the protein level (Extended Data Fig. 1d). These genes were also bound by MYC in the proximity of their transcription start site at the pre-tumoural stage, mostly on canonical E-boxes (CACGTG) (Fig. 1c and Extended Data Fig. 1b). The correlation between MYC and the core snRNP assembly genes also held true in

publicly available data sets from human lymphoma and patients with leukaemia (Extended Data Fig. 1e), and was validated at the RNA level in 29 primary samples (Extended Data Fig. 1f and Supplementary Table 1) and by immunohistochemistry for PRMT5 on 40 samples from large B cell, marginal zone, follicular- and mantle-cell lymphomas (Extended Data Fig. 1g and Supplementary Table 2), confirming the upregulation of PRMT5 in samples overexpressing MYC. Normal tonsils and lymph nodes had lower PRMT5 protein expression (Extended Data Fig. 1g). Importantly, the high expression of *PRMT5* was an accurate predictor of bad prognosis/survival in cohorts of large diffuse B-cell lymphomas (Fig. 1d and Extended Data Fig. 1h). On the basis of these observations, we hypothesized that coordinate increases in the concentrations of core splicing factors, in particular PRMT5, might be critical for sustaining oncogenic growth in MYC-driven tumours. In line with this concept, glioblastoma stem cells, which overexpress MYC, are more sensitive to splicing inhibition than neural stem cells<sup>14</sup>.

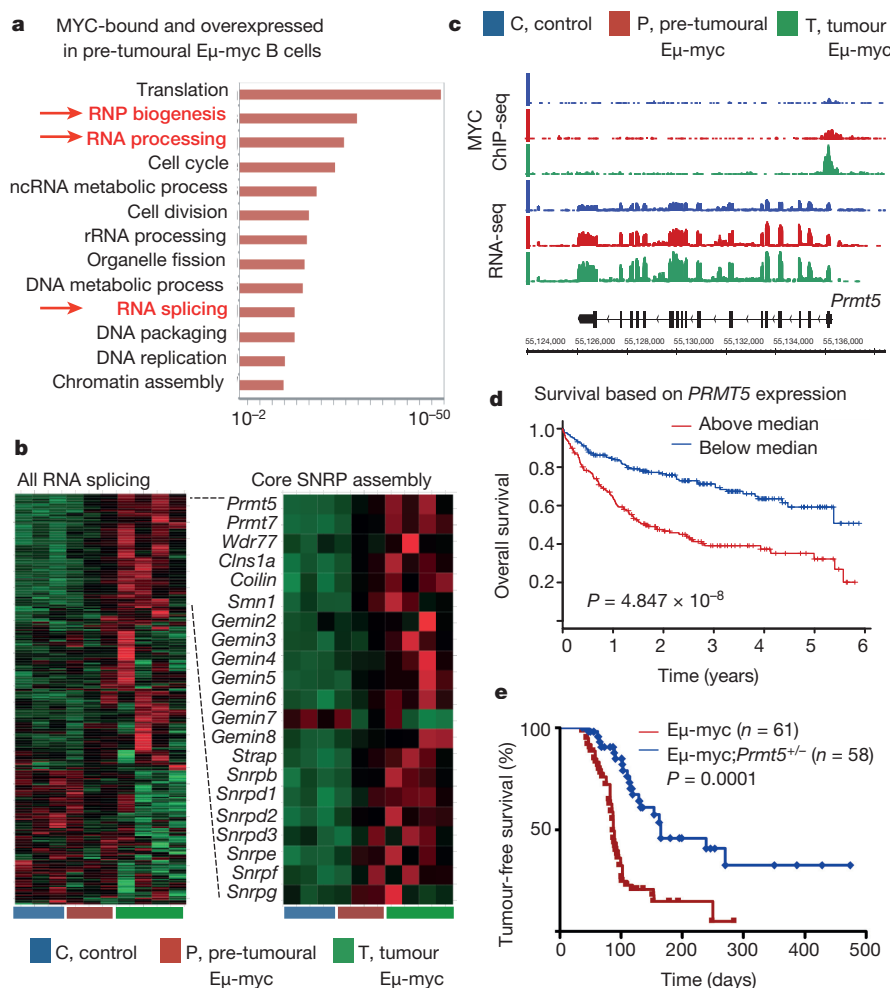
Unlike *Prmt5*<sup>−/−</sup> mice, which are embryonic lethal<sup>15</sup>, *Prmt5*<sup>+/-</sup> mice appear normal and have normal blood counts (data not shown). To address whether PRMT5 is limiting in Myc-induced lymphomagenesis, we thus followed disease development in Eμ-myc;*Prmt5*<sup>+/-</sup> mice. Remarkably, lymphoma development was profoundly delayed in the *Prmt5*<sup>+/-</sup> background (Fig. 1e) ( $P < 0.0001$ ). The Eμ-myc;*Prmt5*<sup>+/-</sup> mice also had significantly reduced disease burden than their age-matched Eμ-myc littermates, both phenotypically and histologically (Extended Data Fig. 2a–c)<sup>16</sup>.

To dissect the mechanistic basis of PRMT5 dependency in MYC-driven lymphomagenesis, we acutely deleted PRMT5 in pre-tumoural bone-marrow-derived Eμ-myc;*Prmt5*<sup>F/F</sup>CreER B cells. Acute *Prmt5* deletion was induced *ex vivo* by the activation of CreER, which significantly reduced viability (Fig. 2a), increasing both apoptosis and G1 arrest (Fig. 2b). Notably, these phenotypic effects of acute *Prmt5* deletion were virtually absent in non-transgenic *Prmt5*<sup>F/F</sup>CreER bone marrow pre-B cells (Fig. 2a, b). Consistent with our previous report<sup>7</sup>, PRMT5 depletion led to a reduction in methylated Sm proteins (Y12; Extended Data Fig. 3a) and the aberrant splicing of *Mdm4* (Extended Data Fig. 3b). We then used RNA-seq to profile the abundance of alternative mRNA species in Eμ-myc;*Prmt5*<sup>F/F</sup>CreER bone marrow pre-B cells in the presence (EtOH) or absence (OHT) of PRMT5. We identified 3,245 differentially expressed genes (Supplementary Table 3). Consistent with the observed phenotype, the functional annotation highlighted defects in cell proliferation, among other pathways (Extended Data Fig. 3c), and we observed that the compiled number of reads in introns was elevated in the absence of PRMT5 (Extended Data Fig. 3d). Interestingly, multivariate analysis of transcript splicing (MATS) identified 153 alternative splicing events (mainly retained introns and skipped exons (Extended Data Fig. 3e and Supplementary Table 4) with weak 5' donor sites) (Extended Data

<sup>1</sup>Institute of Molecular and Cell Biology (IMCB), A\*STAR (Agency for Science, Technology and Research), Singapore 138673, Singapore. <sup>2</sup>Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, 8 Medical Drive, Singapore 117597, Singapore. <sup>3</sup>Cancer Science Institute of Singapore (CSI), National University of Singapore, 14 Medical Drive, Singapore 117599, Singapore. <sup>4</sup>Center for Genomic Science of IIT@SEMM, Fondazione Istituto Italiano di Tecnologia (IIT), Via Adamello 16, 20139 Milan, Italy. <sup>5</sup>Department of Experimental Oncology, European Institute of Oncology (IEO), Via Adamello 16, 20139 Milan, Italy. <sup>6</sup>Institute of High Performance Computing (IHPC), A\*STAR (Agency for Science, Technology and Research), Connexis, Singapore 138632, Singapore.

<sup>7</sup>Bioinformatics Institute (BI), A\*STAR (Agency for Science, Technology and Research), Singapore 138671, Singapore.  
\*These authors contributed equally to this work.





**Figure 1 | MYC directly upregulates the core snRNP assembly genes, including PRMT5.** **a**, Gene set enrichment analysis showing the top enriched categories for Myc-bound and overexpressed transcripts in Eμ-myc pre-tumoural B cells. **b**, Heat map showing the expression of RNA splicing and core SNRP assembly genes in wild-type ( $n = 3$ ), Eμ-myc pre-tumoural ( $n = 3$ ) and Eμ-myc tumour B cells ( $n = 4$ ). **c**, ChIP-seq analysis of MYC binding (top three

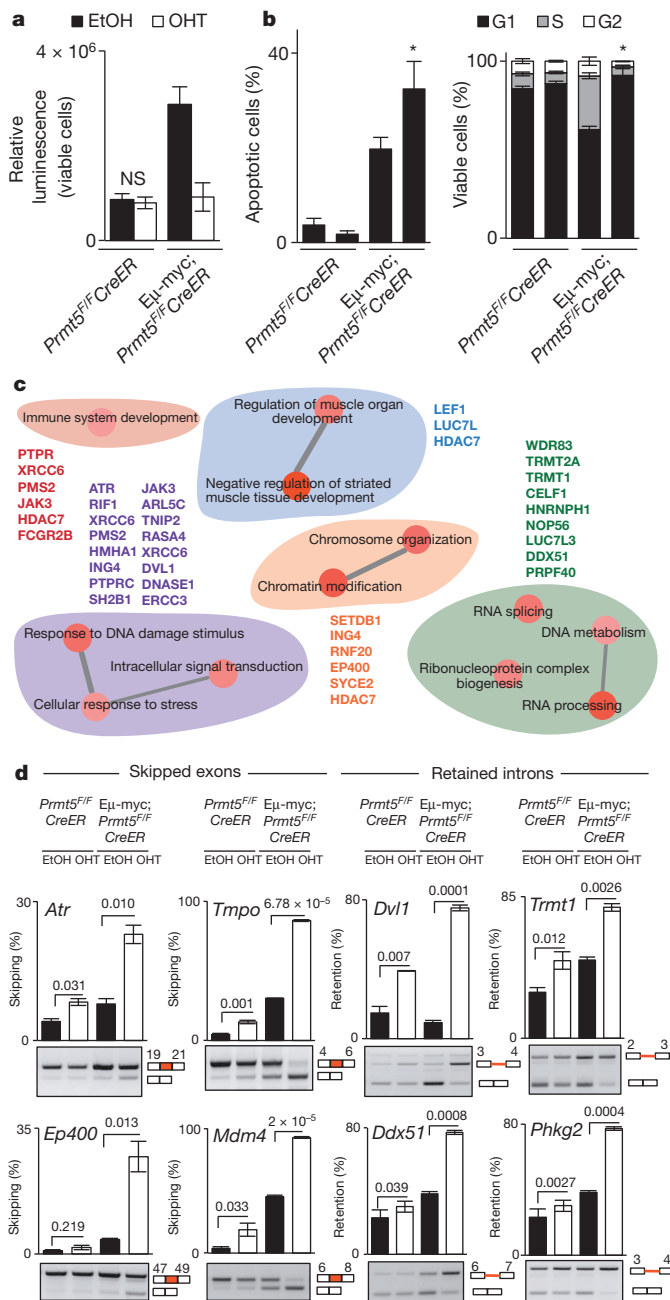
lanes) to the promoter region of *Prmt5* and RNA-seq (bottom lanes) showing the expression of *Prmt5* in control (C), Eμ-myc pre-tumoural (P) and Eμ-myc tumour B cells (T). **d**, Survival of patients with lymphoma stratified by *PRMT5* expression. **e**, Kaplan-Meier analysis of tumour-free survival of Eμ-myc ( $n = 61$ ) and Eμ-myc;*Prmt5*<sup>+/-</sup> mice ( $n = 58$ ).

Fig. 3f), affecting gene products involved in regulating cellular response to stress, signal transduction and chromatin organization (among others) (Fig. 2c). Remarkably, we were able to validate 22 out of 22 of these alternative splicing events in Eμ-myc;*Prmt5*<sup>F/F</sup>CreER bone marrow pre-B cells, while less significant changes were observed in non-transgenic *Prmt5*<sup>F/F</sup>CreER bone marrow pre-B cells (Fig. 2d and Extended Data Fig. 4a). In summary, upon PRMT5 depletion in B cells, we observed a strong correlation between MYC levels and the severity of both splicing defects and cell proliferation (Fig. 2a, b, d). On the basis of our RNA-sequencing results, we predicted that several of the alternative spliced mRNAs, which were generated in cells with reduced PRMT5 levels, would be targeted for nonsense-mediated messenger RNA decay and/or be out of frame (Supplementary Table 5). Accordingly, we verified that the observed alternative splicing events (both skipped exons and retained introns) resulted in the reduction of the full-length proteins in seven cases tested (ATR, MDM4, DVL1, PRPF40, CEP110, PHKG2 and EP400) (Extended Data Fig. 4b).

The critical role of *Prmt5* in lymphoma development suggested that it may also be essential for tumour maintenance. To address this question, we transplanted severe combined immunodeficiency (SCID) recipient mice with tumour cells derived from Eμ-myc;*Prmt5*<sup>F/F</sup>CreER mice, and subsequently deleted *Prmt5* *in vivo* (Fig. 3a). Remarkably, deletion of one copy of PRMT5 slowed disease

onset, while homozygous deletion resulted in full disease-free survival (Fig. 3a and Extended Data Fig. 5a, b). Acute *Prmt5* deletion *ex vivo*, in the same Eμ-myc;*Prmt5*<sup>F/F</sup>CreER lymphoma cells, reduced viability, increased both apoptosis and G1 arrest (Fig. 3b) and induced alternative splicing events similar to those observed in pre-tumoural B cells (Fig. 3c and Extended Data Fig. 5c). Both phenotypic and splicing defects were observed in all samples tested, irrespective of a functional ARF-MDM2-p53 pathway (Extended Data Fig. 5d). The partial reduction of PRMT5 in Eμ-myc;*Prmt5*<sup>+/-</sup>CreER lymphoma cells only resulted in a subset of the observed alternative splicing events (Extended Data Fig. 5e), possibly accounting for the intermediate phenotype observed (Fig. 3a).

PRMT5 depletion in the human Burkitt lymphoma cell lines Raji and Daudi reduced viability (Extended Data Fig. 6a), increased the number of apoptotic cells (Extended Data Fig. 6b, c) and reduced their tumorigenic potential in xenografted mice (Extended Data Fig. 6d). The depletion of PRMT5 also reduced the levels of methylated Sm proteins (Y12) (Extended Data Fig. 6e) and resulted in the aberrant splicing of pre-mRNAs with weak 5'-donor sites, causing both exon skipping and intron retention (Extended Data Fig. 6f), as previously observed in murine cells (Fig. 3c and Extended Data Fig. 5c). In accordance with these data, the depletion of another core component of the splicing machinery (SmB) resulted in the perturbation of similar targets,



**Figure 2 | PRMT5 is essential for maintaining splicing fidelity.** **a**, Viability of  $Prmt5^{F/F} CreER$  and  $E\mu-myc; Prmt5^{F/F} CreER$  pre-B cells after Prmt5 deletion ( $n = 4$  for each genotype). **b**, Apoptosis and the cell-cycle profile of the pre-B cells after Prmt5 deletion, as assessed by flow cytometry ( $n = 5$ ). **c**, Gene ontology of alternatively spliced transcripts detected upon PRMT5 depletion in  $E\mu-myc; Prmt5^{F/F} CreER$  bone marrow B cells. **d**, Validation of alternatively spliced transcripts in  $E\mu-myc; Prmt5^{F/F} CreER$  bone marrow B cells after Prmt5 deletion (OHT) by semi-quantitative PCR. Importantly, these changes were considerably less pronounced after PRMT5 deletion in B cells from  $Prmt5^{F/F} CreER$  mice. Quantification of three independent biological replicates is shown on top, while a representative example is shown in the bottom panel. Data are the average and s.d. Student's *t*-test (two-sided) was used; \* $P < 0.05$  and \*\* $P < 0.01$ .

and in reduced cell viability (Extended Data Fig. 6g). These data confirm that PRMT5 is indeed a key regulator of the core splicing machinery, and is essential for maintaining splicing fidelity in lymphoma. Importantly, in both mouse ( $E\mu-myc$  B cells) and human (Raji/Daudi) lymphoma cells, the downregulation of MYC led to reduced PRMT5, Smd1 and Smd3 mRNA and protein levels (Fig. 3d and

Extended Data Fig. 7a, c), and resulted in aberrant splicing events (Fig. 3e and Extended Data Fig. 7b, d), similar to those observed upon PRMT5 or SmB depletion (Fig. 3c and Extended Data Figs 5c and 6f, g).

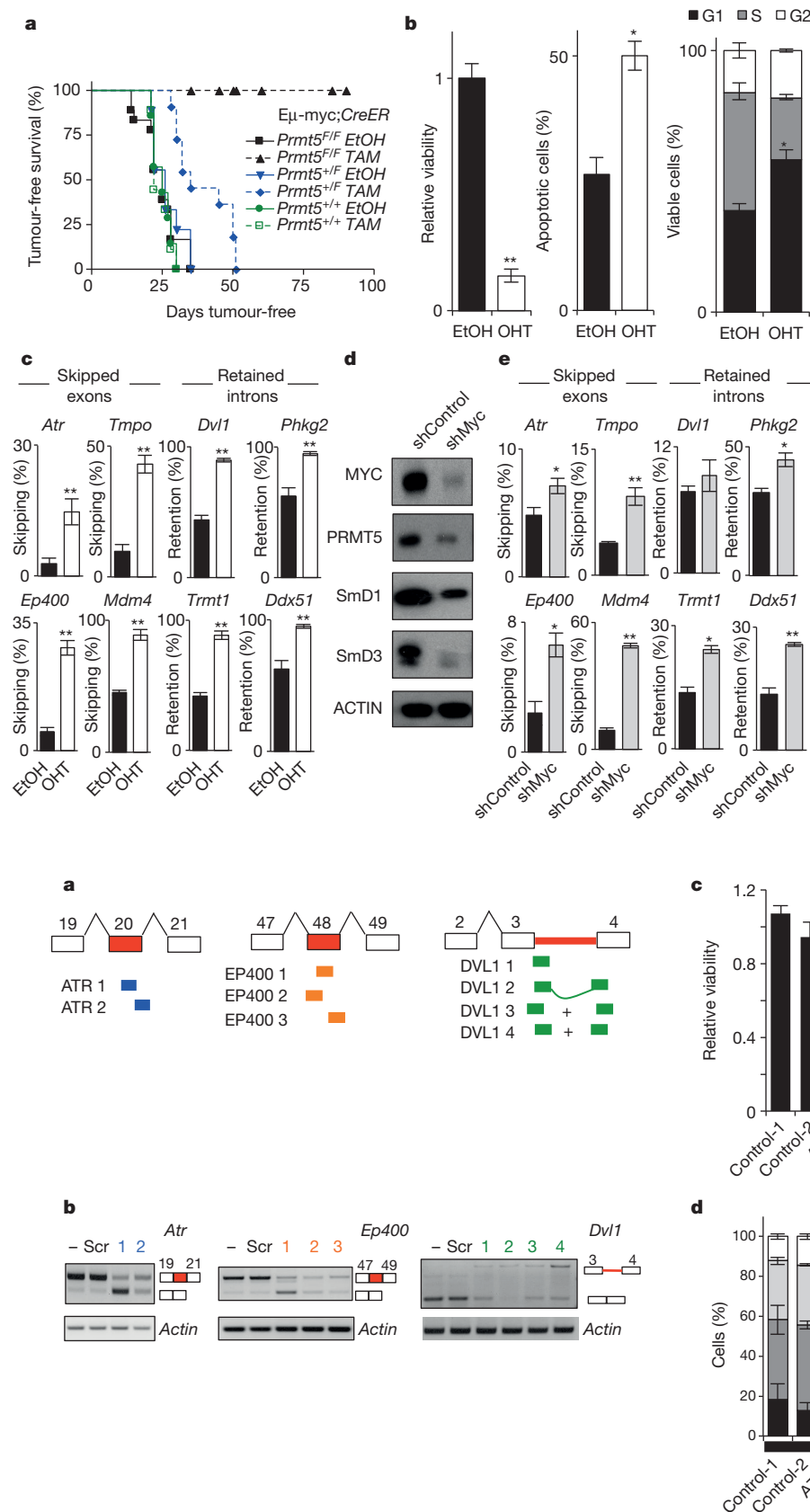
Our data show that the effect of PRMT5 deletion is more pronounced in proliferating cells expressing high MYC levels (Fig. 2a, b, d and Extended Data Fig. 4a), suggesting that systemic *Prmt5* deletion *in vivo* may cause adverse side effects. Indeed, the constitutive deletion of *Prmt5* in adult mice and in embryos during mid-gestation resulted in death<sup>7</sup>. *Prmt5* deletion *in utero* at embryonic day (E)10.5 resulted in a reduction in fetal liver cellularity (Extended Data Fig. 8a), with a reduction of both  $kit^{+}lin^{-}; CD34^{-}$  and  $kit^{+}lin^{-}; CD34^{+}$  cells (Extended Data Fig. 8b). The remaining cells were severely impacted in their ability to form haematopoietic colonies *in vitro* (Extended Data Fig. 8c), and their expression profile indicated a strong upregulation of apoptotic pathways, among others (Extended Data Fig. 8d). We observed similar reductions in cellularity and haematopoietic colony formation in adult  $Prmt5^{F/F} ER$  bone marrow cells after *Prmt5* deletion *in vivo* (Extended Data Fig. 8e, f), which was accompanied by a marked reduction in mature lymphocytes, neutrophils and erythrocytes in the peripheral blood (Extended Data Fig. 8g). Additionally, the deletion of PRMT5 in lethally irradiated mice reconstituted with  $Prmt5^{F/F} ER$  bone marrow haematopoietic progenitors<sup>17</sup> resulted in a significant reduction in bone marrow cellularity, which subsequently led to the death of the animals (Extended Data Fig. 8h).

In summary, the complete elimination of PRMT5 had severe consequences on normal physiology; yet, the overexpression of PRMT5 observed in various haematopoietic malignancies (Fig. 1a–c and Extended Data Fig. 1), the delayed progression of lymphomas after heterozygous deletion of the gene (Fig. 1e), the sensitization by MYC to loss of PRMT5 (Fig. 2a, b) and the dependency on PRMT5 for tumour maintenance (Fig. 3a, b) suggest that a therapeutic window for PRMT5 inhibition may exist.

To further explore the functional consequences of aberrant splicing resulting from PRMT5 depletion, we decided to focus on *Atr* and *Ep400* (skipped exons), given their role in preventing MYC-induced replicative stress<sup>18</sup> and in cooperating with MYC transcriptional activity<sup>19</sup> and MYC-driven cell-cycle progression<sup>20</sup> respectively; and on *Dvl1* (retained intron), given the well-documented interplay between MYC and the Wnt pathway in promoting tumorigenesis<sup>21</sup>. We designed steric hindrance antisense oligonucleotides (ASOs)<sup>22</sup> to induce specific *Atr* exon 20 skipping, *Ep400* exon 48 skipping and *Dvl1* intron 3 retention (Fig. 4a, Supplementary Table 6 and Extended Data Fig. 9a). The electroporation of the ASOs into  $E\mu-myc$  B cells resulted in the specific induction of the alternative splicing event, leading to the reduction of *Atr*, *Ep400* and *Dvl1* full-length mRNAs and protein levels (Fig. 4b and Extended Data Fig. 9b), reduction of cell viability (Fig. 4c) and the induction of apoptosis (Fig. 4d). We have thus identified an essential regulatory mechanism that, through the transcriptional control of PRMT5 and other snRNP components, allows MYC to regulate the maintenance of a fully functional splicing machinery. This enables proper post-transcriptional processing and consequent expression of full-length proteins (for example, ATR, EP400 and DVL1, among others), sustaining cancer cell survival and proliferation. The reversal of a single splicing event (for example, intron retention of *Dvl1*), after PRMT5 depletion, only modestly rescued cell viability (Extended Data Fig. 9c). This is not surprising, given that multiple splicing events are affected upon perturbation of the core splicing machinery.

After our genome-wide mapping of MYC binding and gene regulation<sup>4</sup>, we show here that MYC is a master regulator of the core snRNP machinery, ensuring proper mRNA splicing. Other links had previously been made between MYC and splicing: first, PTPB, hnRNPA1 and hnRNPA2 levels correlate with those of MYC, regulating expression of the PKM2 isoform of pyruvate kinase<sup>23</sup>; second, MYC drives the transcription of the oncogenic splicing regulator, SRSF1 (refs 24, 25).

**Figure 3 | Acute MYC depletion in murine and human lymphoma cells results in aberrant splicing.** **a**, Primary lymphoma cells were isolated from independent tumour-bearing E $\mu$ -myc mice of the indicated genotypes and transplanted into SCID recipients, who were injected intraperitoneally with control (EtOH) or tamoxifen (TAM) 1 week after transplant. The Kaplan–Meier analysis of tumour-free survival of the recipient mice is shown. E $\mu$ -myc;Prmt5<sup>F/F</sup>CreER EtOH ( $n = 21$ ), E $\mu$ -myc;Prmt5<sup>F/F</sup>CreER TAM ( $n = 21$ ), E $\mu$ -myc;Prmt5<sup>+/-</sup>CreER EtOH ( $n = 9$ ), E $\mu$ -myc;Prmt5<sup>+/-</sup>CreER TAM ( $n = 11$ ), E $\mu$ -myc;Prmt5<sup>+/-</sup>CreER EtOH ( $n = 7$ ), E $\mu$ -myc;Prmt5<sup>+/-</sup>CreER TAM ( $n = 9$ ). **b**, Viability (left panel) and apoptosis and cell-cycle profile (middle and right panels) of the pre-B cells after Prmt5 deletion ( $n = 5$ ). **c**, Validation of alternatively spliced transcripts (retained introns and skipped exons) after Prmt5 deletion (OHT) in E $\mu$ -myc lymphoma cells ( $n = 5$ ). **d**, MYC, Prmt5, SmD1, SmD3 and  $\beta$ -actin protein expression in whole-cell lysates from E $\mu$ -myc B cells infected with viruses encoding non-targeting short hairpin RNA (shRNA) (shControl) and or shRNAs targeting MYC (shMyc). **e**, Validation of alternatively spliced transcripts after MYC knockdown in E $\mu$ -myc B cells ( $n = 4$ ). Data are the average and s.d. Student's  $t$ -test (two-sided) was used; \* $P < 0.05$ , \*\* $P < 0.01$ .



**Figure 4 | Antisense oligonucleotides targeting ATR-, EP400- and DVL1-alternative splicing mimic the cell-cycle arrest/apoptotic response induced by MYC/PRMT5 depletion.** **a**, Schematic representation of ASOs designed to induce exon skipping (ATR and EP400) and intron retention (DVL1). **b**, Validation of efficacies of ASOs in inducing alternatively spliced transcripts. Representative gel images are shown ( $n = 4$ ). **c**, Cell viability of E $\mu$ -myc B cells

after electroporation with the respective ASOs ( $n = 4$ ). **d**, Cell-cycle profiles of E $\mu$ -myc B cells after electroporation with the respective ASOs ( $n = 4$ ). The  $P$  values indicate the significance in the difference in percentage apoptosis, compared with controls. Data are the average and s.d. Student's  $t$ -test (two-sided) was used; \* $P < 0.05$ , \*\* $P < 0.01$ .



Our report, however, provides the first evidence that MYC regulates the constitutive splicing machinery, including the Sm proteins and PRMT5 (refs 6, 7). Taken together, our results support a model (Extended Data Fig. 10) whereby MYC overexpression forces the cell to rely on high levels of PRMT5 and mature snRNPs to sustain splicing fidelity, most probably because of the documented direct and indirect RNA amplification levels in cancer cells<sup>3,4</sup> (see Supplementary Information for additional discussion). These data point at the possibility of drugging the spliceosome, either by directly affecting its core components<sup>12</sup> or by targeting PRMT5 methyltransferase activity, as a potentially viable strategy in MYC-driven tumours, although potential side effects (Extended Data Fig. 8) will have to be carefully taken into account.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 10 January 2014; accepted 24 February 2015.

Published online 11 May 2015.

- Dang, C. & McMahon, S. B. Emerging concepts in the analysis of transcriptional targets of the MYC oncoprotein: are the targets targetable? *Genes Cancer* **1**, 560–567 (2010).
- Guccione, E. *et al.* Myc-binding-site recognition in the human genome is determined by chromatin context. *Nature Cell Biol.* **8**, 764–770 (2006).
- Lin, C. Y. *et al.* Transcriptional amplification in tumor cells with elevated c-Myc. *Cell* **151**, 56–67 (2012).
- Sabo, A. *et al.* Selective transcriptional regulation by Myc in cellular growth control and lymphomagenesis. *Nature* **511**, 488–492 (2014).
- Walz, S. *et al.* Activation and repression by oncogenic MYC shape tumour-specific gene expression profiles. *Nature* **511**, 483–487 (2014).
- Meister, G. *et al.* Methylation of Sm proteins by a complex containing PRMT5 and the putative U snRNP assembly factor pICln. *Curr. Biol.* **11**, 1990–1994 (2001).
- Bezzi, M. *et al.* Regulation of constitutive and alternative splicing by PRMT5 reveals a role for Mdm4 pre-mRNA in sensing defects in the spliceosomal machinery. *Genes Dev.* **27**, 1903–1916 (2013).
- Barna, M. *et al.* Suppression of Myc oncogenic activity by ribosomal protein haploinsufficiency. *Nature* **456**, 971–975 (2008).
- Papaemmanuil, E. *et al.* Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. *N. Engl. J. Med.* **365**, 1384–1395 (2011).
- Yoshida, K. *et al.* Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* **478**, 64–69 (2011).
- Damm, F. *et al.* Mutations affecting mRNA splicing define distinct clinical phenotypes and correlate with patient outcome in myelodysplastic syndromes. *Blood* **119**, 3211–3218 (2012).
- Bonnal, S., Vigevani, L. & Valcarcel, J. The spliceosome as a target of novel antitumour drugs. *Nature Rev. Drug Discov.* **11**, 847–859 (2012).
- Friesen, W. J. *et al.* The methylosome, a 20S complex containing JBP1 and pICln, produces dimethylarginine-modified Sm proteins. *Mol. Cell. Biol.* **21**, 8289–8300 (2001).
- Hubert, C. G. *et al.* Genome-wide RNAi screens in human brain tumor isolates reveal a novel viability requirement for PHF5A. *Genes Dev.* **27**, 1032–1045 (2013).
- Tee, W. W. *et al.* Prmt5 is essential for early mouse development and acts in the cytoplasm to maintain ES cell pluripotency. *Genes Dev.* **24**, 2772–2777 (2010).
- Harris, A. W. *et al.* The E mu-myc transgenic mouse. A model for high-incidence spontaneous lymphoma and leukemia of early B cells. *J. Exp. Med.* **167**, 353–371 (1988).
- Hemann, M. T. *et al.* Evasion of the p53 tumour surveillance network by tumour-derived MYC mutants. *Nature* **436**, 807–811 (2005).
- Murga, M. *et al.* Exploiting oncogene-induced replicative stress for the selective killing of Myc-driven tumors. *Nature Struct. Mol. Biol.* **18**, 1331–1335 (2011).
- Frank, S. R. *et al.* MYC recruits the TIP60 histone acetyltransferase complex to chromatin. *EMBO Rep.* **4**, 575–580 (2003).
- Fujii, T., Ueda, T., Nagata, S. & Fukunaga, R. Essential role of p400/mDomino chromatin-remodeling ATPase in bone marrow hematopoiesis and cell-cycle progression. *J. Biol. Chem.* **285**, 30214–30223 (2010).
- Ge, X. & Wang, X. Role of Wnt canonical pathway in hematological malignancies. *J. Hematol. Oncol.* **3**, 33 (2010).
- Pramono, Z. A., Yee, W. C., Lai, P. S. & Wee, K. B. Antisense oligonucleotides and uses thereof. WO Patent WO/2011/078,797.
- David, C. J., Chen, M., Assanah, M., Canoll, P. & Manley, J. L. HnRNP proteins controlled by c-Myc deregulate pyruvate kinase mRNA splicing in cancer. *Nature* **463**, 364–368 (2010).
- Anczukow, O. *et al.* The splicing factor SRSF1 regulates apoptosis and proliferation to promote mammary epithelial cell transformation. *Nature Struct. Mol. Biol.* **19**, 220–228 (2012).
- Das, S., Anczukow, O., Akerman, M. & Krainer, A. R. Oncogenic splicing factor SRSF1 is a critical transcriptional target of MYC. *Cell Rep.* **1**, 110–117 (2012).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank S. Campaner, M. Pelizzola, D. Messerschmidt and P. R. Kaldas for sharing protocols and for discussions; F. B. M. Ibrahim and V. S. Y. Tay for help with histopathology work; BRC Shared Facilities for technical support; D. Bararia for help in the fluorescence-activated cell sorting analysis. We are grateful to X. Ruan, W. Soon and the GIS Genome Sequencing Team for help with the Solexa high-throughput sequencing, and to the entire E.G. laboratory for discussion. This work was supported by an AGA-SINGA (Singapore Graduate Award) fellowship to M.B. and by IMCB, A-STAR. E.G. acknowledges support from JCO-ASTAR grants 1134c001 and 11/03/FG/07/04. B.A. acknowledges support from the European Research Council, the Italian health ministry and the Italian Association for Cancer Research (AIRC).

**Author Contributions** C.M.K., M.B., E.G.: overall design of the project. C.M.K., M.B., W.X.A., S.X.T., A.S., F.P.H.G.: acquisition of samples. C.M.K., M.B., E.G., D.L.: detailed molecular and genomics approaches. D.L.: bioinformatic analysis. C.M.K., M.B., BA, E.G.: organization of experiments and figures. M.B., M.O.: fluorescence-activated cell sorting analysis. K.B.W.: antisense oligonucleotide design. M.A.H., T.S.Y.: histopathology. C.M.K. and E.G. wrote the manuscript with comments from all authors.

**Author Information** RNA-seq and microarray datasets generated in this study have been deposited in Gene Expression Omnibus database in the Superseries GSE61638. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.G. ([eguccione@imcb.a-star.edu.sg](mailto:eguccione@imcb.a-star.edu.sg)).

## METHODS

No statistical methods were used to predetermine sample size.

**Mouse strains and genotyping.** The floxed PRMT5 mice were described previously<sup>7</sup>. 4-Hydroxytamoxifen (4-OHT)-inducible conditional knockouts were created by crossing *Prmt5<sup>F/F</sup>* mice with Rosa26-CreERT2 transgenic mice<sup>26</sup> (in mixed C57BL/6 × 129S1/SvImJ background). The Eμ-myc were purchased from the Jackson Laboratory and SCID mice were obtained from BRC, A\*STAR.

**Animal studies.** The experimental protocol was approved by the Institutional Animal Care and Use Committee (IACUC), and the animals were maintained under compliance with institutional guidelines. The sample size was not pre-selected and no inclusion/exclusion criteria were used. All mice were monitored daily for signs of morbidity and lymphoma development. Whole blood was collected from the tail vein of mice using heparinized microhaematocrit capillary tubes (Fisher 22-362-566) and analysed using an NK MEK-6318 haematology analyser (Nihon Koden Corporation). For tissue collection, the mice were killed according to IACUC guidelines, and the spleens, tumours and bone marrow were collected for subsequent analysis.

**Isolation of haematopoietic cell populations.** Haematopoietic cells were extracted from fetal liver. Flow cytometric analyses by FACSaria (BD Biosciences) were performed using a standard method. Briefly, blocking non-specific binding in mouse serum was followed by antibody staining on ice for 30 min. After washing, cells were suspended in 0.5 ml of phosphate-buffered saline with propidium iodide or Hoechst 33258 for dead-cell discrimination. On the screen of flow cytometry, homogenous populations in cell size (forward scatter-side scatter window) and viable cells were gated and analysed for individual antigen expressions. Lineage-markers included Gr1, Ter119, CD3, CD4, CD8 and B220. Monoclonal antibodies were purchased from BD Biosciences or eBioscience: PE-conjugated anti-Gr1 (RB6-8C5), Ter119 (TER-119), CD3 (145-2C11), CD4 (RM4-5), CD8 (53-6.7), B220 (RA3-6B2), PE-Cy7-conjugated anti-c-Kit (2B8) and FITC-conjugated CD34 (RAM34).

For colony formation assays, haematopoietic cells were extracted from E14.5 fetal liver or adult bone marrow. Methocult M3630 and M3434 (Stem Cell Technologies) assays were set up according to the manufacturer's protocol. Colonies were enumerated 10–14 days after plating under bright field microscope.

**Immunohistochemistry staining.** Automated immunohistochemistry (IHC) staining and counterstaining were performed on a Leica Bond-™ autostainer.

**Cell culture.** Primary Eμ-myc lymphoma and bone marrow cells were isolated and maintained, as previously described<sup>27,28</sup>. Raji and Daudi cells were maintained in RPMI1640 and 293T cells were maintained in DMEM. All media were supplemented with 10% FBS and 1% penicillin–streptomycin. All cells were grown in a humidified incubator at 37 °C and 5% CO<sub>2</sub>. The shRNA sequences for lentiviral-based knockdown were as follows: shPRMT5-1, CCGGGGCTCAAGCCACCAA TCTATGCTCGAGCATAGATTGGTGGCTTGAGCCTTTTGTG; shPRMT5-2, CCGGCCATCCTCTTCCCTATTAAGCTCGAGCTTAATAGGGAAGAGGA TGGGTTTTTGTG; shSMB-1, CCGGCCACAAGGAAGAGGTAAGTCTCTCGAG AACATACCTCTTCTTGTGTTTTTGTG; shSMB-2, CCGGCACATGAATTT GATCCTCTGTCTCGAGACAGAGATCAAATTCATGTGTTTTTGTG; shMYC (mouse), GGAGATGATGACCGAGTTA; shMYC-1 (human), GATGAGGAAG AAATCGATG; shMYC-2 (human), GATGAGGAAGAAATCGATA.

**Antibodies.** The following antibodies were used: actin (Santa Cruz, SC-47778), PRMT5 (Santa Cruz, SC-22132), c-MYC (Santa Cruz, SC-40), SNRPD1 (Abcam, ab50940), SNRPD3 (Abcam, ab111094), and Smith antigen (Abcam, ab3138).

**ChIP sequencing and RNA sequencing library preparation.** ChIP-seq and RNA-seq were performed according to standard methods (Illumina), or as previously described<sup>8</sup>.

**RNA sequencing and analysis.** Paired-end sequencing (150 base pairs) was performed on an Illumina Nextseq 500. Reads were mapped to the mm9 mouse genome assembly using Tophat (<http://tophat.cbcb.umd.edu/>) (version 2.0.9) with the following parameters: [-p 8-splice-mismatches 1-segment-mismatches 2]. Aligned reads were then quantified for expression using the Cufflinks suite version 2.1.1 (<http://cufflinks.cbcb.umd.edu/>) and edgeR (<http://www.bioconductor.org/packages/release/bioc/html/edgeR.html>). Downstream manipulation of RNA-seq results was done with CummeRbund (version 2.0) and with in-house scripts. Genes were considered to be significantly differentially expressed at  $P < 0.05$  and  $\text{abs}(\log_2[\text{FPKM}])/\text{ratio} > 1$ . Gene set testing used the mroast function from the R package limma. To determine differential splicing events, MATS 3.0.8 beta (<http://rnaseq-mats.sourceforge.net/>) was used to count junction reads and reads falling into the tested region within ENSEMBL gene definitions. Splicing events were labelled significant if the sum of the reads supporting a specific event exceeded ten reads, and  $P < 0.05$ .

**ChIP-sequencing.** ChIP-seq data sets analysed were obtained from ref. 4. Alignments were generated from Illumina Fastq files to the mm9 genome using

Bowtie 0.12.8 (<http://bowtie-bio.sourceforge.net/index.shtml>). Peak calling was done using MACS (<https://github.com/taoliu/MACS/>) (version 2.0.9) with the following parameters: [-bw 300 -m 5,30-shiftsize 100 -p 1e-3-slocal = 1000 -B-keep-dup = 1]. Peak distributions were defined using RefSeq annotation. Distances within -5 kilobases (kb)/+2 kb of a gene transcription start site were defined to be the promoter. To identify overlapping peaks between data sets, a region was first defined as a  $\pm 1$  kb window from an enriched peak centre. This region was then compared against the list of similarly defined regions from a corresponding data set.

**Other bioinformatics analysis.** All GO term analysis used DAVID (<http://david.abcc.ncifcrf.gov/>)<sup>29</sup>. Density plots of genomic loci were generated using IGB (<http://bioviz.org/igb/>). GO reduction and connectivity graphs were done with REVIGO (<http://revigo.irb.hr/>) and Cytoscape version 3.0.2 ([www.cytoscape.org/](http://www.cytoscape.org/)). Public Gene Expression Omnibus (GEO) data sets were retrieved using the R package 'GEOquery'. Survival analysis was done in R using the 'survival' package.

**Accession numbers.** Microarray data from public data sets were obtained from GEO with the accession numbers GSE4475, GSE22470 and GSE10846. RNA-sequencing data from ref. 4 were obtained from GSE51011. RNA-sequencing and microarray data sets generated in this study have been deposited in Gene Expression Omnibus database in the Superseries GSE61638.

**Tamoxifen injections in vivo and 4-OHT treatment in vitro.** For *in vivo* experiments, 2 mg tamoxifen (Sigma, T5648) was administered intraperitoneally in adult mice for three consecutive days. For *in vitro* experiments, cells were treated with 50 nM 4-OHT or an equal volume of EtOH as a control for 24 h, after which the cells were washed and resuspended in fresh media. In all experiments with *Prmt5<sup>F/F</sup>* CreER-derived cells or animals, we used ROSA26:CreER counterparts as negative controls, ensuring that the addition of OHT or tamoxifen was not toxic<sup>30</sup>. The doses of tamoxifen and 4-OHT were carefully titrated to the minimum necessary for effective recombination, with minimal side effects.

**Cell viability assays.** Cell viability was assessed by a CellTiter-Glo Luminescent Cell Viability Assay (Promega), according to the manufacturer's protocol.

**Flow cytometry analysis.** The spleens and lymphomas were collected and single-cell suspensions were isolated in PBS with 10% FBS. The cells were stained with 500  $\mu\text{g ml}^{-1}$  propidium iodide in 1% w/v Triton X-100 in PBS with 2 mg RNase A (Sigma) at room temperature (23 °C) for 30 min, and subsequently analysed by flow cytometry analysis. Dead cells and debris were excluded by gating on the basis of their forward and side scatter characteristics.

**Lymphoma transplant studies.** These were done as previously described<sup>31</sup>. Briefly, 6- to 8-week old SCID recipient mice were transplanted with  $10^6$  primary lymphoma cells by tail vein injection and monitored daily for signs of disease manifestation. Recipient mice from different litters were randomized equally for group allocation. No blinding was done.

**Statistical analysis.** Student's *t*-test (two-sided) was applied, and changes were considered statistically significant when  $P < 0.05$ . In the figures, changes with  $*P < 0.05$  and  $**P < 0.01$ . The data were normally distributed and variation within and between groups was not estimated. The sample size was not pre-selected and no inclusion/exclusion criteria were used. The data shown are the averages and s.d. of at least three biological replicates (that is, cells isolated from at least three independent mice). Statistical analysis used Microsoft Excel or GraphPad Prism software.

**ASOs.** Novel ASOs were applied to bind to nascent transcripts of a target gene via Watson–Crick bonding to exert steric hindrance effect against splicing factors to modulate splicing. The ASOs were rationally designed for optimal efficiency in inducing splicing modulation, as previously described<sup>32,33</sup>. Briefly, ASO target sites were selected by a computational algorithm that accounted for co-transcriptional binding accessibilities, binding thermodynamics and presence of regulatory splicing motifs. All the designed ASOs were synthesized by Sigma Aldrich (Singapore) as single-stranded 2'-O-methyl modified RNA bases linked with phosphorothioate backbone.

Each ASO inducing specific exon-skipping events in *ATR* and *EP400* genes was designed to bind exonic sequences to mask respective splicing motifs that were required for the proper splicing of the target exon. On the other hand, ASOs inducing a specific intron retention event in the *DVL1* gene were designed to bind the target intron splice sites either separately or simultaneously, as depicted in Fig. 4a, to mask the boundaries of its flanking exons. In the latter, either two distinct ASOs or a dual-targeting ASO (that is, an ASO targeting two non-consecutive target sites<sup>34</sup>) were used.

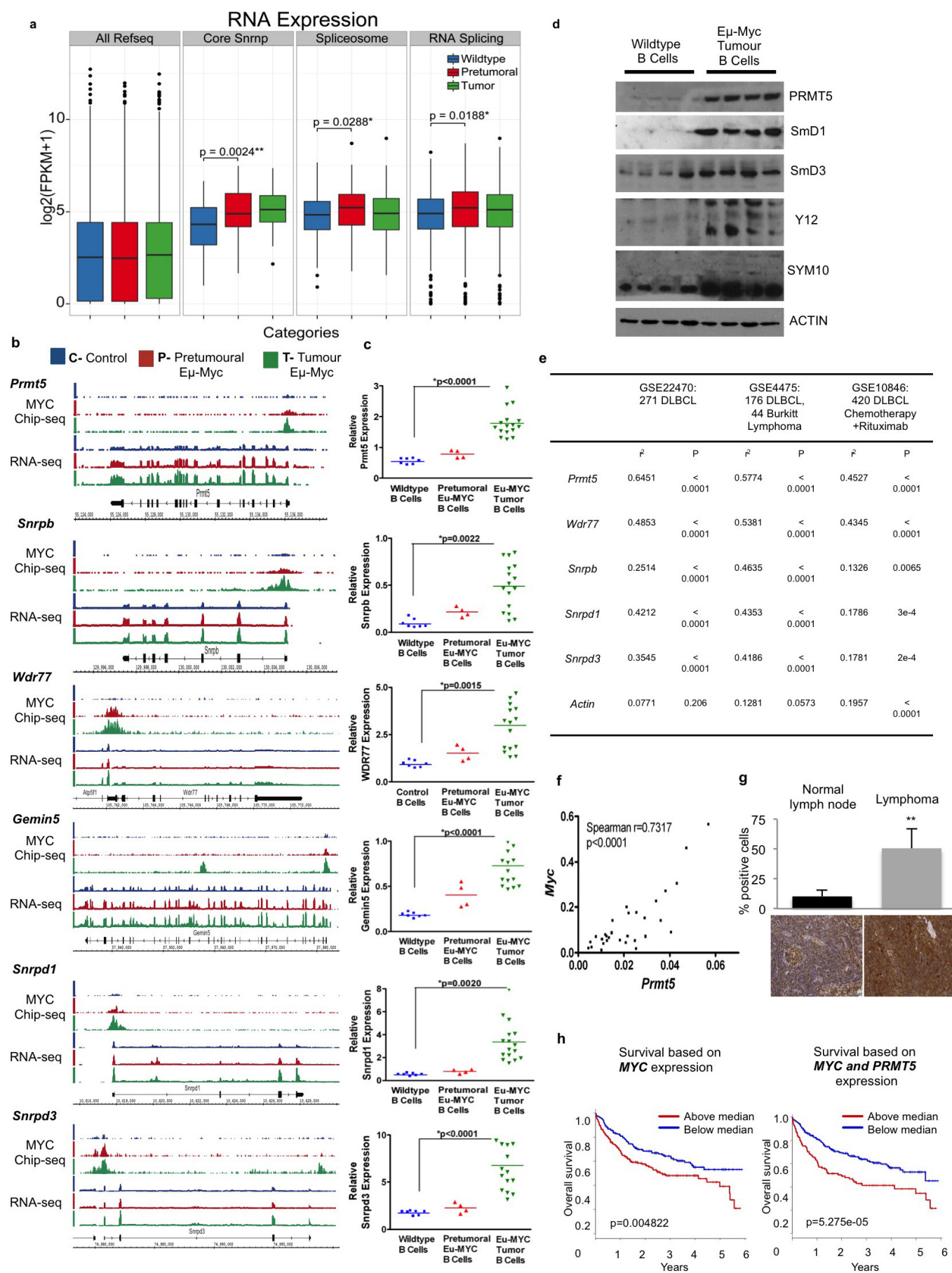
The sequence of each ASO and the experimental design for validating *DVL1* targeting ASOs is tabulated in Supplementary Table 6.

**ASO transfection protocol.** ASOs were electroporated into Eμ-myc B cells at a final concentration of 100 nM using the Neon Transfection System (Invitrogen). The electroporation parameters were optimized individually for B cells isolated

from separate mice, and were typically  $1750\text{ V} \times 20\text{ s} \times \text{one pulse}$ ,  $1350\text{ V} \times 30\text{ s} \times \text{one pulse}$  or  $1200\text{ V} \times 20\text{ s} \times \text{two pulses}$ .

26. Hameyer, D. *et al.* Toxicity of ligand-dependent Cre recombinases and generation of a conditional Cre deleter mouse allowing mosaic recombination in peripheral tissues. *Physiol. Genom.* **31**, 32–41 (2007).
27. Eischen, C. M., Weber, J. D., Roussel, M. F., Sherr, C. J. & Cleveland, J. L. Disruption of the ARF-Mdm2-p53 tumor suppressor pathway in Myc-induced lymphomagenesis. *Genes Dev.* **13**, 2658–2669 (1999).
28. Schmitt, C. A. & Lowe, S. W. Bcl-2 mediates chemoresistance in matched pairs of primary E(mu)-myc lymphomas in vivo. *Blood Cells Mol. Dis.* **27**, 206–216 (2001).
29. Huang D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4**, 44–57 (2009).
30. Schmidt-Suppran, M. & Rajewsky, K. Vagaries of conditional gene targeting. *Nature Immunol.* **8**, 665–668 (2007).
31. Schmitt, C. A. & Lowe, S. W. Bcl-2 mediates chemoresistance in matched pairs of primary E(mu)-myc lymphomas in vivo. *Blood Cells Mol. Dis.* **27**, 206–216 (2001).
32. Wee, K. B. *et al.* Dynamics of co-transcriptional pre-mRNA folding influences the induction of dystrophin exon skipping by antisense oligonucleotides. *PLoS One* **3**, e1844 (2008).
33. Pramono, Z. A. *et al.* A prospective study in the rational design of efficient antisense oligonucleotides for exon skipping in the DMD gene. *Hum. Gene Ther.* **23**, 781–790 (2012).
34. Pao, P. W., Wee, K. B., Yee, W. C. & Dwipramono, Z. A. Dual masking of specific negative splicing regulatory elements resulted in maximal exon 7 inclusion of SMN2 gene. *Mol. Ther.* **22**, 854–861 (2014).





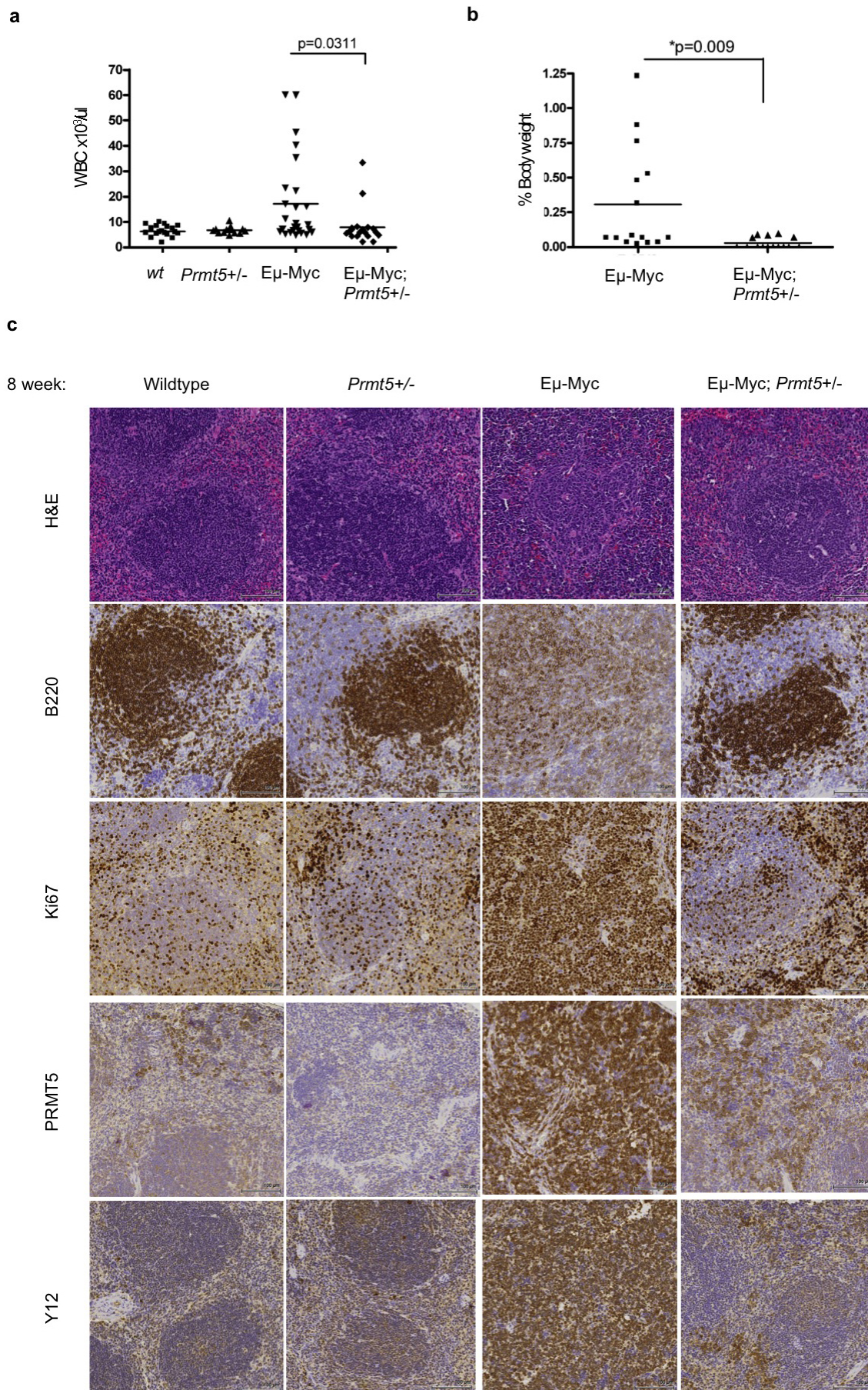
**Extended Data Figure 1 | MYC directly upregulates the core snRNP**

**assembly genes.** **a**, RNA-seq expression of all Refseq genes, core Snrnp genes, spliceosome genes and RNA splicing genes in wild-type, Eμ-myc pre-tumoural B cells and Eμ-myc tumour B cells. **b**, ChIP-seq analysis of MYC binding (top lanes) and RNA-seq (bottom lanes) showing the expression of 'core snRNPs assembly' genes in control (C), Eμ-myc pre-tumoural (P) and Eμ-myc tumours (T). Student's *t*-test (two-tailed) was used; \**P* < 0.05, \*\**P* < 0.01.

**c**, Quantitative PCR validation of the expression of representative 'core snRNPs assembly' genes in wild-type B cells (*n* = 7), Eμ-myc pre-tumoural B cells (*n* = 4) and Eμ-myc tumour cells (*n* = 16). **d**, Immunoblots of representative

'core snRNPs assembly' proteins in wild-type B cells (*n* = 4) and Eμ-myc tumour cells (*n* = 4). **e**, Spearman's  $r^2$  and *P* values for the correlation between the expression of *MYC* and *PRMT5*, *WDR77*, *SNRPB*, *SNRPD1* and *SNRPD3* in publicly available lymphoma data sets. **f**, Correlation between the expression of *PRMT5* and *MYC* in 29 samples from patients with primary leukaemia and lymphoma. **g**, Quantification of PRMT5 staining in normal human lymph nodes and lymphoma. A representative image for IHC staining of PRMT5 in normal human tonsil and follicular lymphoma. **h**, Survival of patients with lymphoma stratified for *PRMT5* and *MYC* expression, individually and combined.

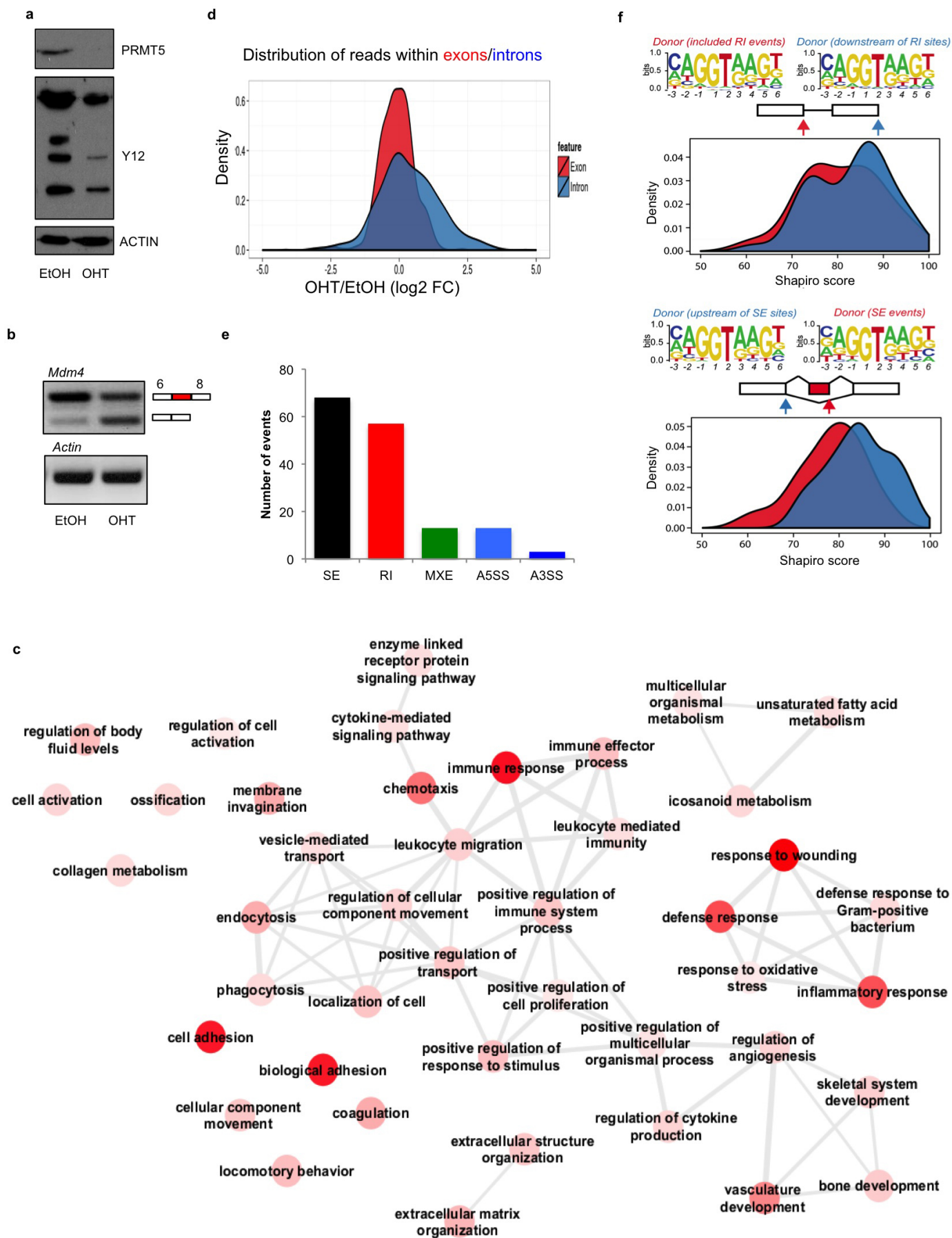






**Extended Data Figure 2 | Reduced disease burden in Eμ-myc;Prmt5<sup>+/-</sup> mice.** **a**, Analysis of WBC counts in 8-week-old wild-type ( $n = 20$ ), Prmt5<sup>+/-</sup> ( $n = 13$ ), Eμ-myc ( $n = 21$ ) and Eμ-myc;Prmt5<sup>+/-</sup> ( $n = 19$ ) littermates. Each point represents one animal. **b**, Analysis of tumour burden in 8-week-old

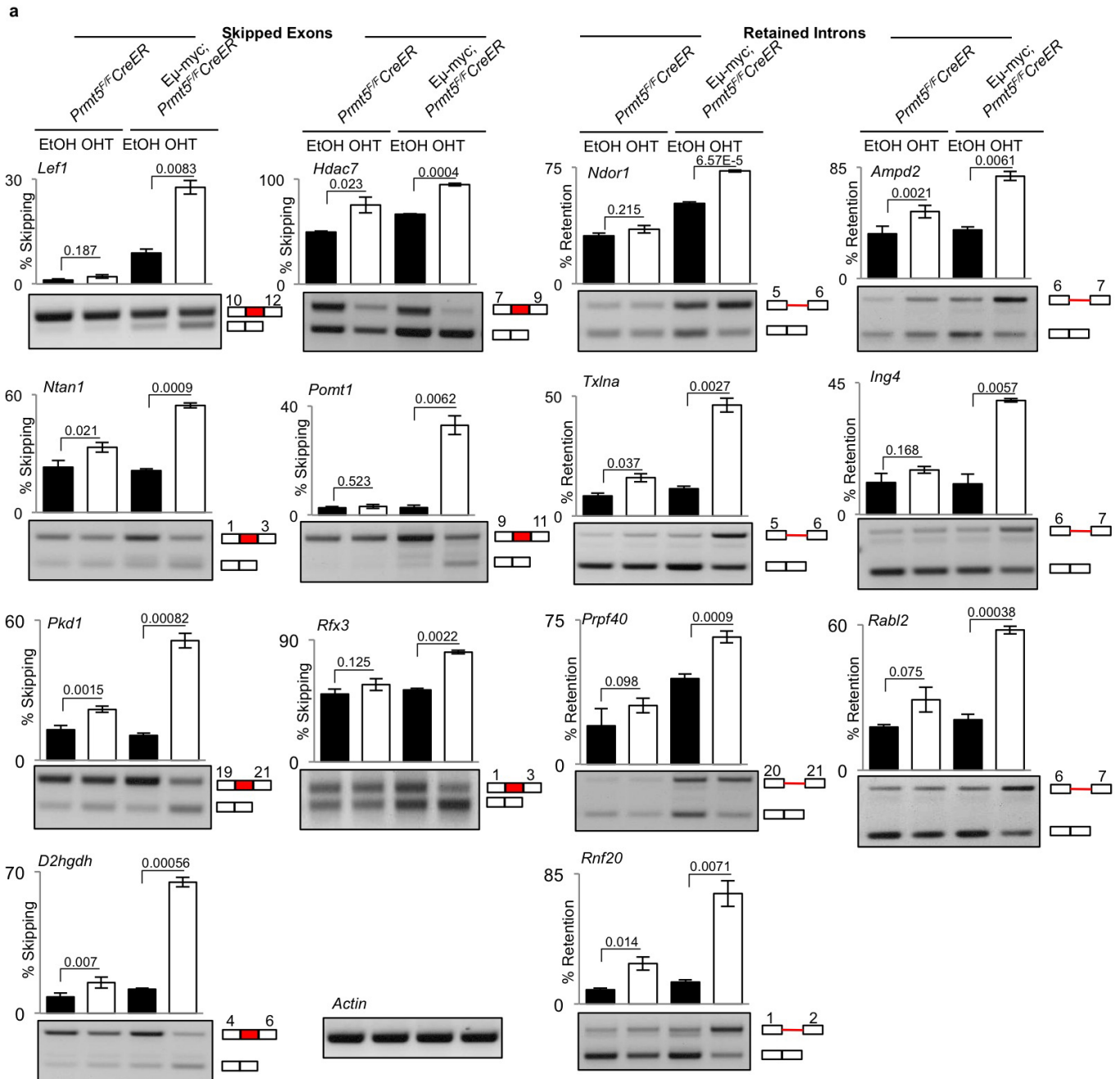
Eμ-myc ( $n = 15$ ) and Eμ-myc;Prmt5<sup>+/-</sup> ( $n = 15$ ) littermates. Each point represents one animal. **c**, Representative images showing haematoxylin and eosin and IHC for PRMT5, B220, Ki67 and Y12 in the spleens of 8-week-old, age-matched wild-type, Prmt5<sup>+/-</sup>, Eμ-myc and Eμ-myc;Prmt5<sup>+/-</sup> littermates.



**Extended Data Figure 3 | PRMT5 depletion leads to a reduction in functioning of the core splicing machinery.** **a**, A representative immunoblot showing the reduction of PRMT5 protein levels after OHT treatment, and a corresponding decrease in Y12 levels (methylated Sm proteins). **b**, Validation of alternative splicing of Mdm4 mRNA in E $\mu$ -myc B cells after PRMT5 deletion. **c**, Functional annotation of genes affected by alternative splicing events (either exon skipping or intron retention). **d**, Distribution of reads within exons and introns (ratio of OHT/EtOH). **e**, MATS output:

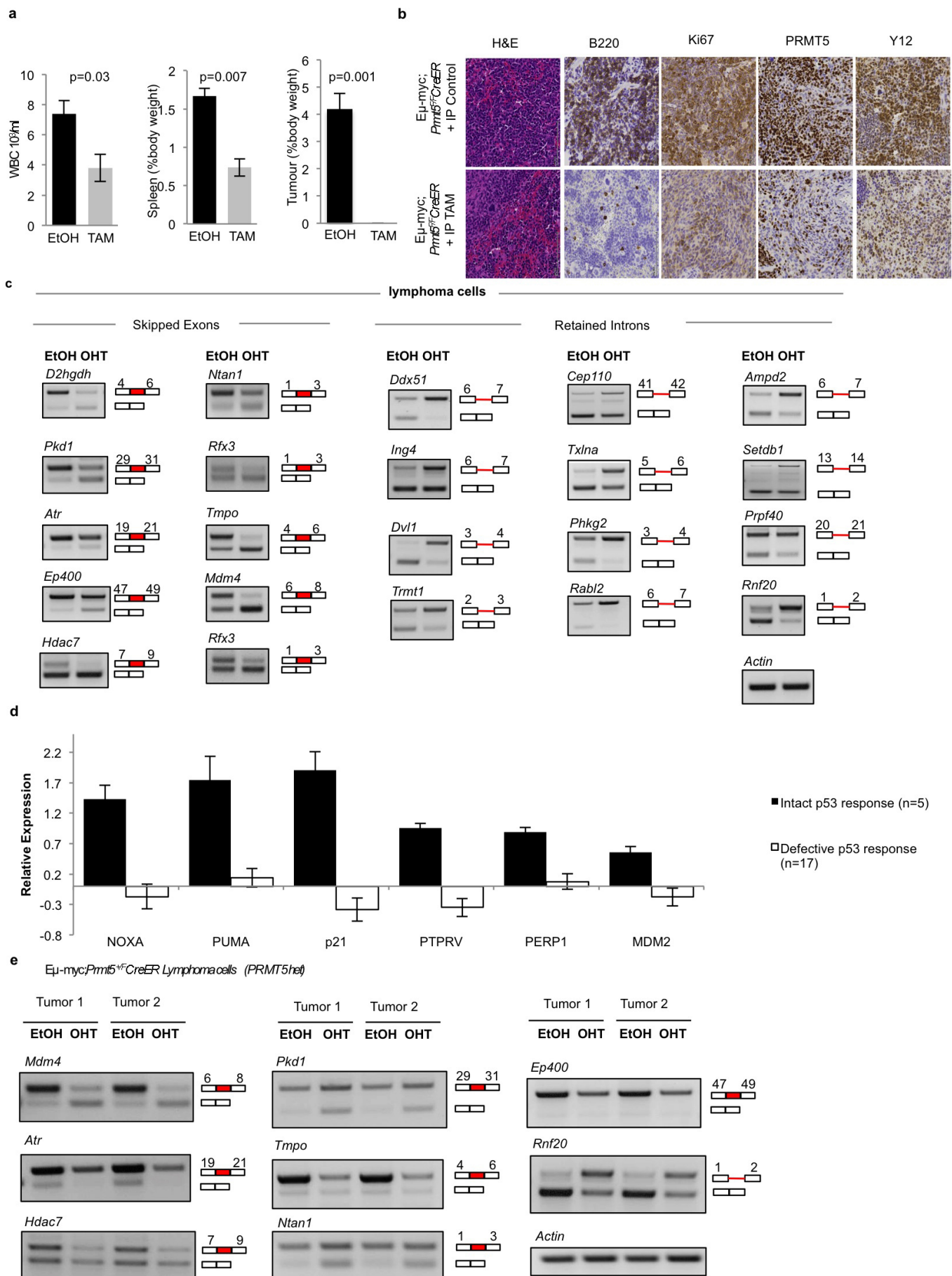
quantification of skipped exons (SE), retained introns (RI), mutually exclusive exons (MXE), A5SS and A3SS (alternative 5' or 3' splice site). **f**, Shapiro score of the 5' donor sites of retained intron events identified by MATS. A smooth density estimate is drawn as calculated by a Gaussian kernel. Top: sequence logo of the 5' donor of retained intron events (left) compared with the 5' donor of the downstream exon (right), detected in PRMT5-depleted cells. Bottom: sequence logo of the 5' donor sites of the skipped exon events (right), compared with the 5' donor of the upstream exon (left).





**Extended Data Figure 4 | Alternative splicing events in Eμ-myc;*Prmt5*<sup>F/F</sup>*CreER* bone marrow pre-B cells lead to reduction in full length protein levels.** **a**, Validation of additional alternatively spliced transcripts in Eμ-myc;*Prmt5*<sup>F/F</sup>*CreER* and *Prmt5*<sup>F/F</sup>*CreER* bone marrow B cells after *Prmt5* deletion (OHT) by semi-quantitative PCR. Quantification of three

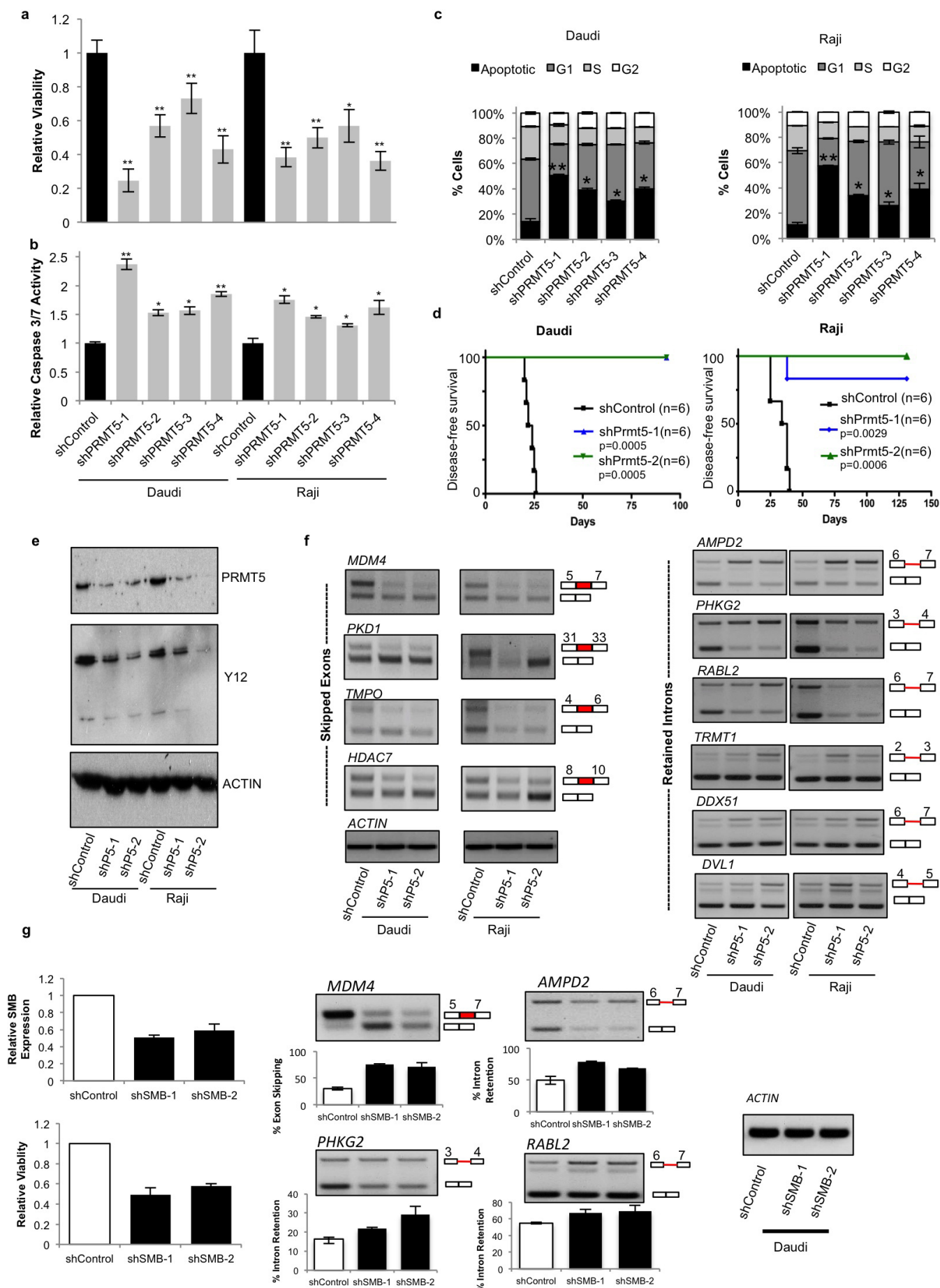
independent biological replicates is shown on top, while a representative example is shown in the bottom panel. **b**, Quantification of selected proteins by immunoblotting or flow cytometry in Eμ-myc;*Prmt5*<sup>F/F</sup>*CreER* bone marrow B cells after *Prmt5* deletion (OHT).





**Extended Data Figure 5 | Alternative splicing events in Eμ-myc;Prmt5<sup>F/F</sup> CreER lymphoma cells.** **a**, Disease burden of recipient mice ( $n = 5$  for each group), as assessed by white blood cell (WBC) counts (left panel), spleen weight (middle panel) and tumour weight (right panel), 3 weeks after transplantation. **b**, Representative images showing the haematoxylin and eosin and IHC for B220, Ki67, PRMT5 and Y12 in the spleens of recipient mice. **c**, Validation of additional alternatively spliced transcripts in Eμ-myc;Prmt5<sup>F/F</sup> CreER lymphoma B cells after Prmt5 deletion (OHT) by semi-quantitative PCR. Representative gel images are shown ( $n = 5$ ). **d**, Expression of p53 target genes after Prmt5 deletion in Eμ-myc;Prmt5<sup>F/F</sup> CreER lymphoma cells, as a demonstration of functional/inactive p53 response. Of the 22 lymphomas, each

isolated from independent tumour-bearing Eμ-myc;Prmt5<sup>F/F</sup> CreER mice, five (22.72%) showed the upregulation of classical p53 target genes in response to PRMT5 deletion (indicating a functional p53 pathway), while 17 (77.27%) did not (indicating an inactive p53 pathway), rates that are similar to previous reports<sup>26</sup>. **e**, Validation of alternatively spliced transcripts in Eμ-myc;Prmt5<sup>F/F</sup> CreER lymphoma B cells after Prmt5 deletion (OHT) by semi-quantitative PCR. Representative gel images are shown ( $n = 3$ ). **f**, Validation of additional alternatively spliced transcripts in after Myc knockdown in Eμ-myc lymphoma B cells after Prmt5 deletion (OHT) by semi-quantitative PCR. Representative gel images are shown ( $n = 4$ ).

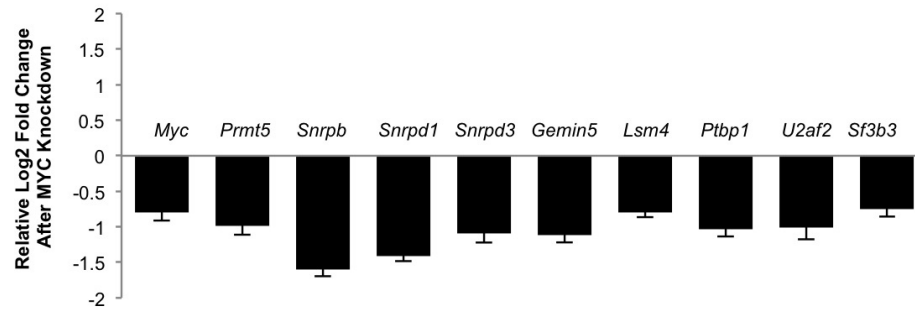


**Extended Data Figure 6 | PRMT5 and SmB depletions in human Burkitt lymphoma lines lead to reduced viability and splicing defects.** **a**, Relative viability of Raji and Daudi cells upon PRMT5 depletion with four independent shRNAs (shPrmt5-1 to shPRMT5-4) ( $n = 4$ ). **b**, Apoptosis profile of Raji and Daudi cells after PRMT5 knockdown (shPrmt5-1 to shPRMT5-4) ( $n = 4$ ). **c**, Cell-cycle profile of Raji and Daudi cells after PRMT5 knockdown (shPrmt5-1 and shPRMT5-4) ( $n = 4$ ). **d**, Control and PRMT5 depleted (shPrmt5-1 and shPRMT5-2) Daudi and Raji cells were xenografted into SCID recipients. The Kaplan–Meier analysis of tumour-free survival of the recipient mice is shown;  $n$ , number of recipient mice analysed. **e**, PRMT5 and Y12 (methylated Sm

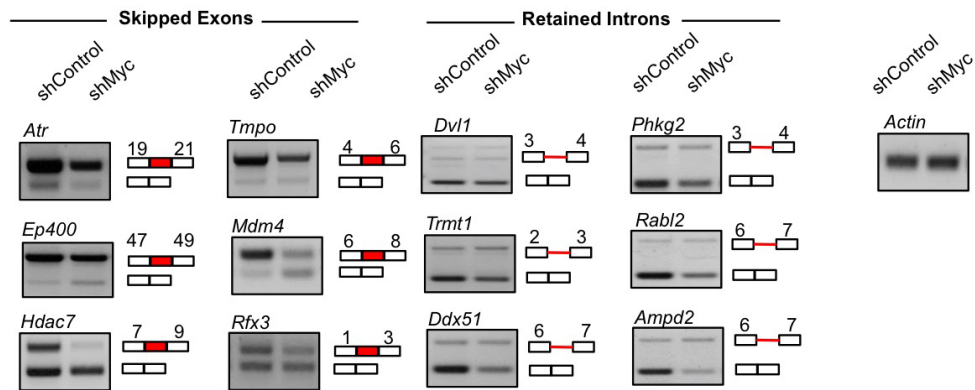
proteins) levels upon PRMT5 depletion in Raji and Daudi cells. A representative blot is shown. **f**, Validation of additional alternatively spliced transcripts (retained introns and skipped exons) after PRMT5 knockdown in Daudi and Raji cells by semi-quantitative PCR ( $n = 4$ ). **g**, Validation of SMB knockdown in Daudi cells (upper left panel) and relative viability after SMB knockdown (lower left panel) ( $n = 3$ ). Validation of additional alternatively spliced transcripts (retained introns and skipped exons) after SMB knockdown in Daudi cells by semi-quantitative PCR and quantification of gels by Image J. The data presented in this figure are the average and s.d. Student's  $t$ -test (two sided) was used;  $*P < 0.05$ ,  $**P < 0.01$ .



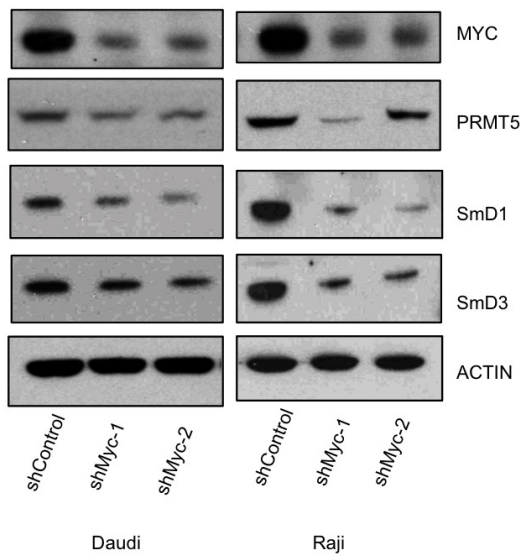
**a**



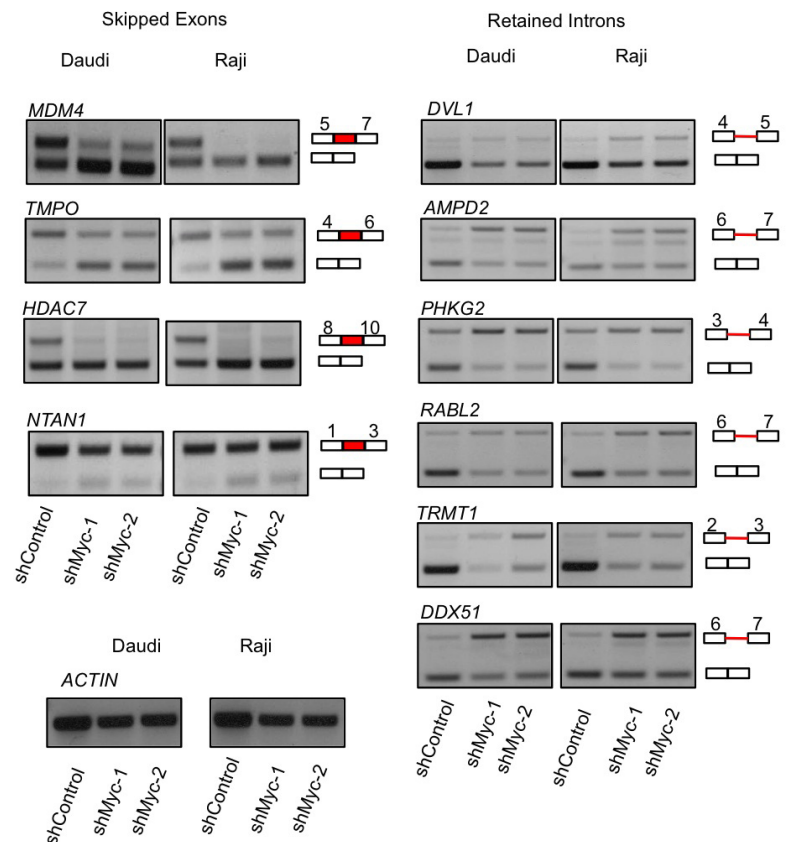
**b**



**c**

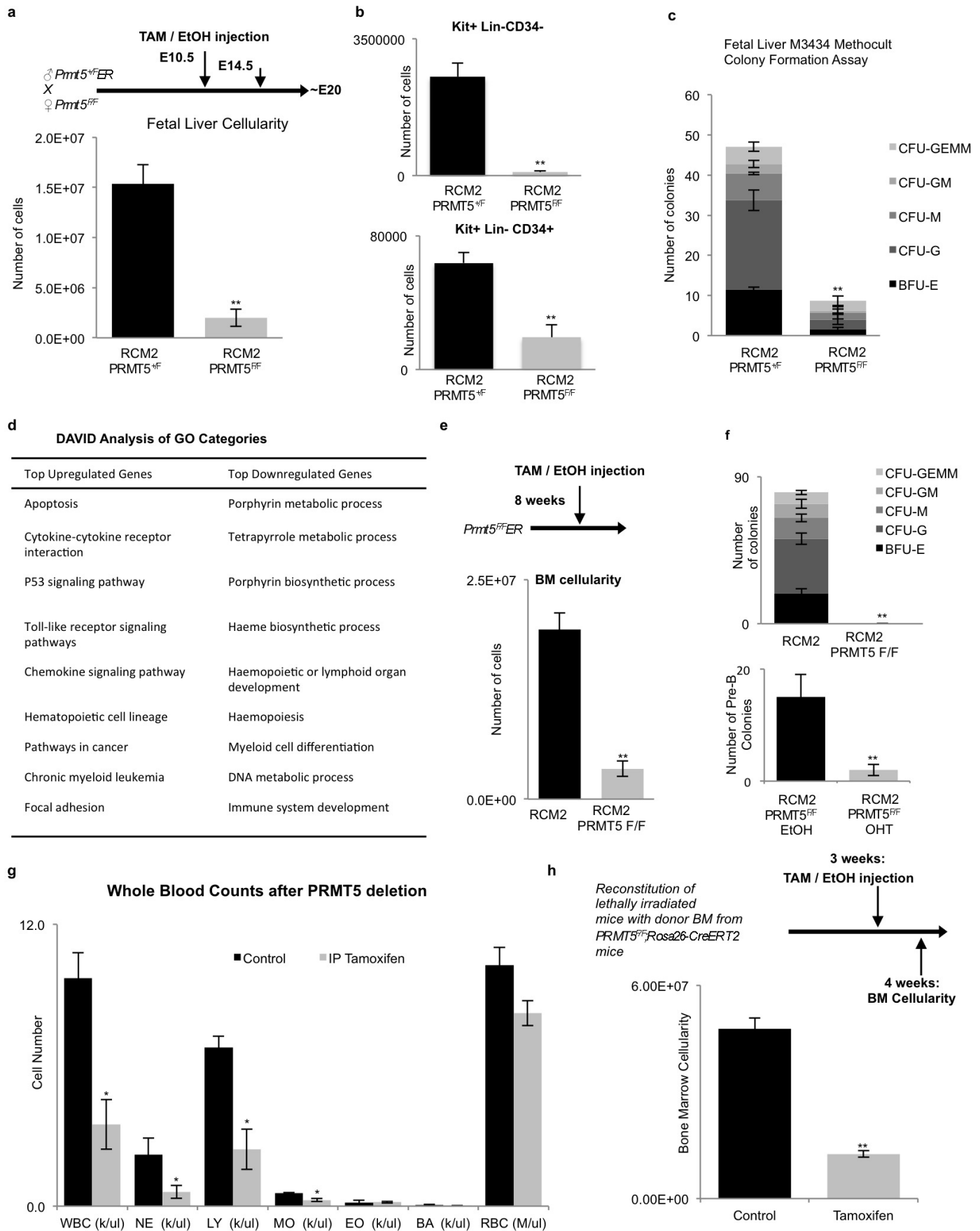


**d**



**Extended Data Figure 7 | Myc depletion leads to splicing defects both in mouse and in human lymphoma cells.** **a.** Relative expression of selected 'core snRNPs assembly' genes following MYC knockdown in E $\mu$ -myc B cells, assessed by quantitative real-time PCR ( $n = 3$ ). **b.** Validation of alternatively spliced transcripts (retained introns and skipped exons) after MYC knockdown in E $\mu$ -myc B cells by semi-quantitative PCR ( $n = 3$ ; representative gel images are shown). **c.** MYC, PRMT5, SmD1, SmD3 and  $\beta$ -actin protein expression in whole cell lysates from Daudi and Raji cells infected with viruses encoding non-

targeting shRNA (shControl) and two independent sequences targeting MYC (shMyc-1, shMyc-2). **d.** Validation of alternatively spliced transcripts (retained introns and skipped exons) after MYC knockdown in Daudi and Raji cells by semi-quantitative PCR ( $n = 3$ ). shControl, scramble control shRNA; shMyc-1, shMyc-2, shRNA sequences targeting MYC. The data presented in this figure are the average and s.d. Student's  $t$ -test (two-sided) was used; \* $P < 0.05$ , \*\* $P < 0.01$ .

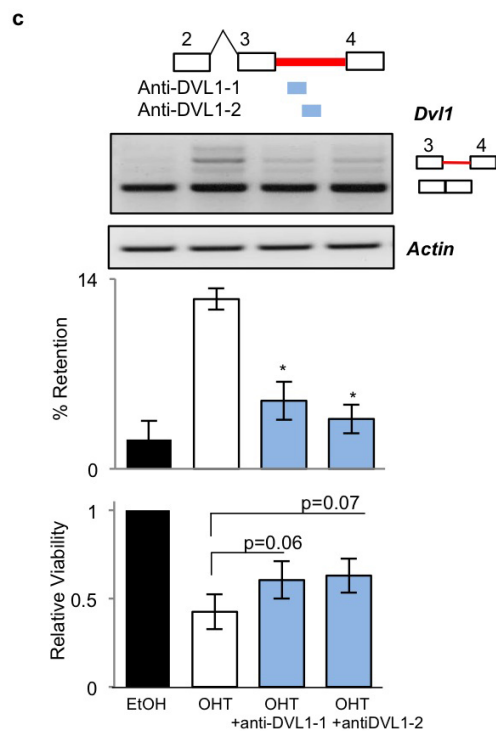
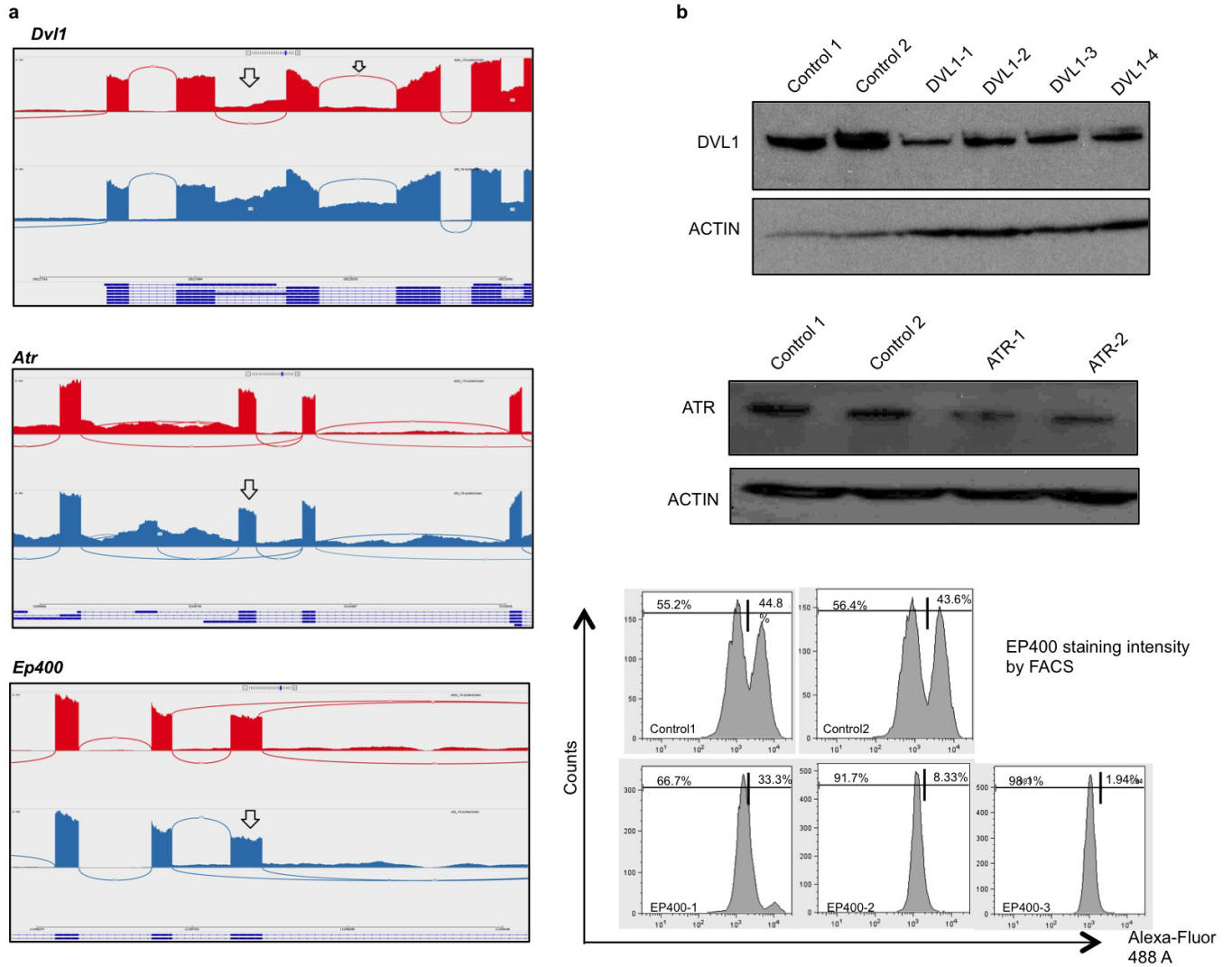




**Extended Data Figure 8 | PRMT5 is required for normal haematopoiesis.**

**a**, Fetal liver cellularity at E14.5, after tamoxifen injection at E10.5 (*Prmt5*<sup>+/F</sup>*CreER*; *n* = 3, *Prmt5*<sup>F/F</sup>*CreER*; *n* = 5). **b**, Flow cytometry analysis of Kit<sup>+</sup>Lin<sup>−</sup>CD34<sup>+</sup> (bottom panel), Kit<sup>+</sup>Lin<sup>−</sup>CD34<sup>−</sup> (top panel), using fetal liver cells from **a**. Total number of cells is indicated (*Prmt5*<sup>+/F</sup>*CreER*; *n* = 3, *Prmt5*<sup>F/F</sup>*CreER*; *n* = 5). **c**, Methocult M3434 colony formation assay using fetal liver cells from **a**. CFU-GEMM, colony-forming unit—granulocyte, erythrocyte, monocyte/macrophage, megakaryocyte; CFU-GM, colony-forming unit—granulocyte, macrophage; CFU-M, colony-forming unit—macrophage; CFU-G, colony-forming unit—granulocyte; BFU-E, burst-forming

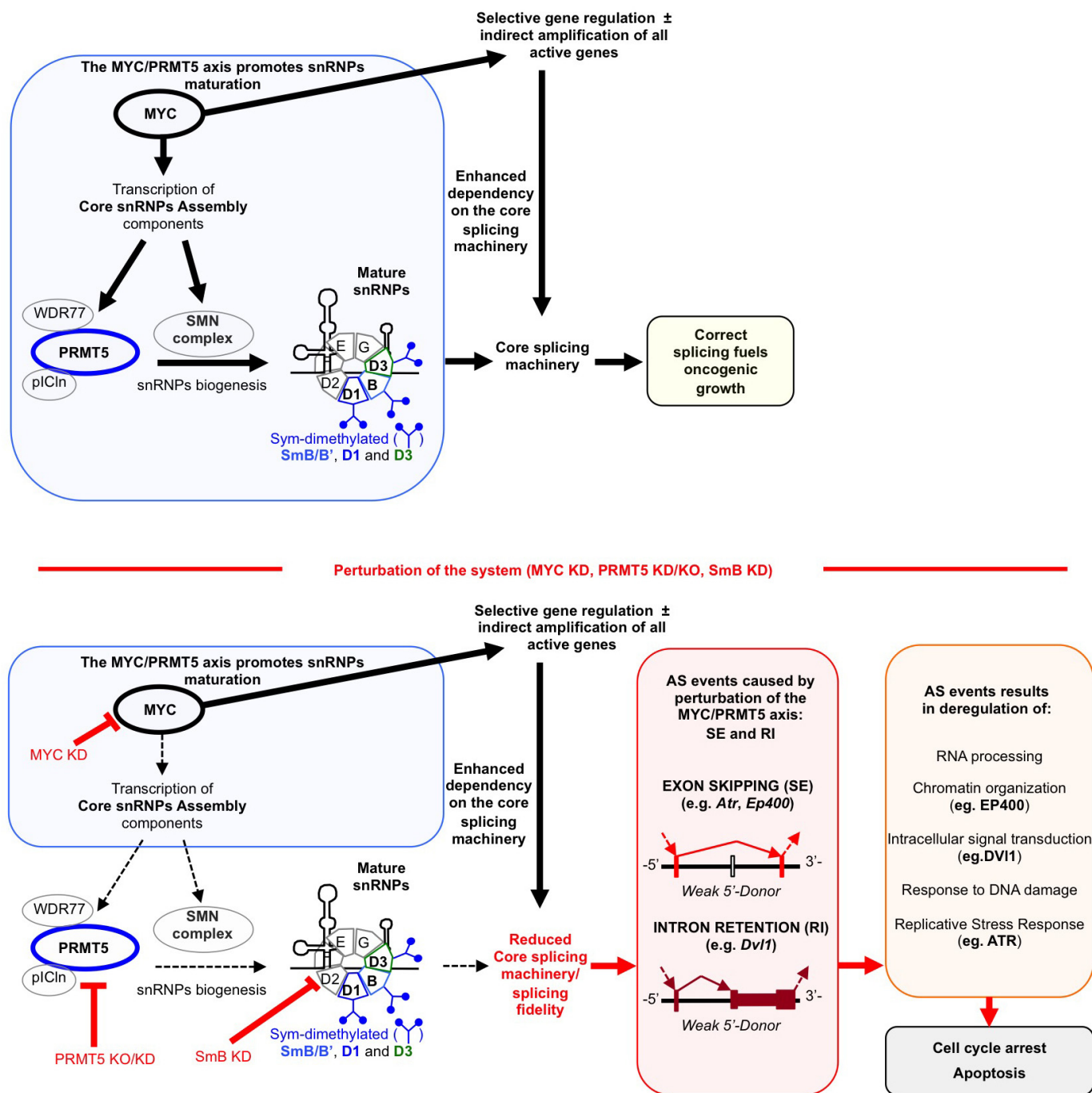
unit—erythrocyte. **d**, GO analysis of differentially expressed genes after PRMT5 deletion in fetal liver cells (*n* = 4). **e**, Bone marrow cellularity of adult mice after tamoxifen injection (*n* = 3 in each group). **f**, Methocult M3434 colony formation assay using bone marrow cells from **e**. **g**, Whole blood counts (WBC) 5 days after PRMT5 deletion *in vivo* (*n* = 5). NE, neutrophils; LY, lymphocytes; MO, monocytes; EO, eosinophils; BA, basophils; RBC, red blood cells. **h**, Bone marrow cellularity after selective deletion of PRMT5 in the haematopoietic system (*n* = 4 mice in each group). The data presented in this figure are the average and s.d. Student's *t*-test (two-sided) was used; \**P* < 0.05, \*\**P* < 0.01.



**Extended Data Figure 9 | Antisense oligonucleotides targeting ATR, EP400 and DVL1 lead to the reduction of their full-length protein levels.** **a**, Sashimi plots showing alternatively spliced transcripts of DVL1, ATR and EP400 after PRMT5 depletion in E $\mu$ -myc;*Prmt5*<sup>F/F</sup>*CreER* cells. **b**, Quantification of protein expression after electroporation of E $\mu$ -myc B cells with the respective ASOs ( $n = 3$ ). **c**, Upper panel: schematic representation of ASOs designed to block the

intron retention in *Dvl1* induced by PRMT5 knockout. Middle panel: validation of efficacies of ASOs in reversing the alternative splicing of *Dvl1*, after PRMT5 knockout ( $n = 3$ ). Bottom panel: cell viability of E $\mu$ -myc B cells 2 days after the electroporation with the respective ASOs. The data presented in this figure are the average and s.d. Student's *t*-test (two-sided) was used; \* $P < 0.05$ , \*\* $P < 0.01$ .





**Extended Data Figure 10 | Graphical summary.** Top panel: black lines indicate MYC direct transcriptional upregulation of *PRMT5* and other components of the core snRNP assembly machinery, which ensures splicing fidelity. Bottom panel: red arrows indicate the perturbation of the

MYC-PRMT5 axis, which leads to a reduction in splicing fidelity within the cell, skipped exons and retained introns of genes, such as *Ep400*, *Dvl1* and *Atr* (which harbour exons with weak 5'-donor sites), downregulation of their protein levels and, consequently, cell-cycle arrest and apoptosis.

# Cytosolic extensions directly regulate a rhomboid protease by modulating substrate gating

Rosanna P. Baker<sup>1</sup> & Siniša Urban<sup>1</sup>

Intramembrane proteases catalyse the signal-generating step of various cell signalling pathways, and continue to be implicated in diseases ranging from malaria infection to Parkinsonian neurodegeneration<sup>1–3</sup>. Despite playing such decisive roles, it remains unclear whether or how these membrane-immersed enzymes might be regulated directly. To address this limitation, here we focus on intramembrane proteases containing domains known to exert regulatory functions in other contexts, and characterize a rhomboid protease that harbours calcium-binding EF-hands. We find calcium potently stimulates proteolysis by endogenous rhomboid-4 in *Drosophila* cells, and, remarkably, when rhomboid-4 is purified and reconstituted in liposomes. Interestingly, deleting the amino-terminal EF-hands activates proteolysis prematurely, while residues in cytoplasmic loops connecting distal transmembrane segments mediate calcium stimulation. Rhomboid regulation is not orchestrated by either dimerization or substrate interactions. Instead, calcium increases catalytic rate by promoting substrate gating. Substrates with cleavage sites outside the membrane can be cleaved but lose the capacity to be regulated. These observations indicate substrate gating is not an essential step in catalysis, but instead evolved as a mechanism for regulating proteolysis inside the membrane. Moreover, these insights provide new approaches for studying rhomboid functions by investigating upstream inputs that trigger proteolysis.

Cell membranes are both controlled borders with the outside world as well as dynamic platforms for organizing cell signalling, metabolic pathways, and ultrastructure assembly. All of these key events rely on enzymes that reside directly within the cell membrane, yet achieving a mechanistic understanding of how these specialized enzymes function within this environment has proved challenging.

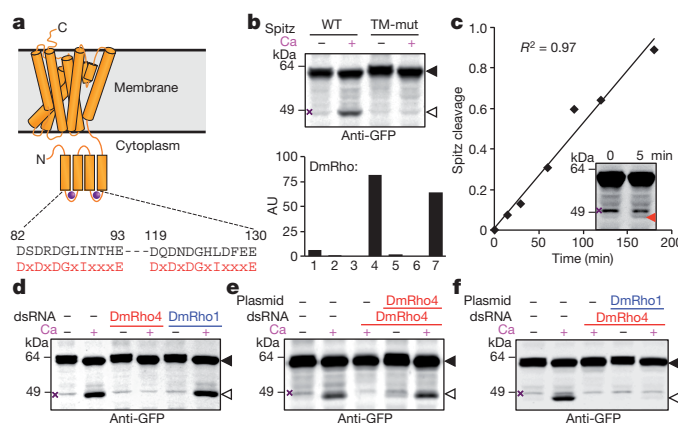
Intramembrane proteases catalyse the committed, signal-generating step of several key signalling pathways by cleaving transmembrane proteins within the membrane<sup>1–3</sup>. Their importance is emphasized by repeated implication in disease.  $\gamma$ -Secretase generates the amyloid- $\beta$  peptide in Alzheimer's disease<sup>4,5</sup>, but more recently has been successfully targeted in a spectrum of cancers<sup>6</sup>, because its activating cleavage of the Notch receptor triggers signalling<sup>2</sup>. Site-2 protease family metalloenzymes liberate transcription factors from the membrane to control cholesterol and fatty-acid composition of membranes<sup>1</sup>, and signalling circuits that control virulence in pathogenic bacteria<sup>7</sup>. Rhomboid serine proteases are a family of master regulators that initiate epidermal growth factor signalling during *Drosophila* development<sup>3,8</sup>, but more recently have been implicated in cleaving adhesins during malaria invasion<sup>9</sup>, and regulating mitochondrial quality control to guard against Parkinson's disease<sup>10</sup>.

Since peptide bond cleavage is irreversible in the cell, precise regulation of protease activity is paramount. Yet it is generally thought that intramembrane proteases are constitutively active enzymes over which the cell cannot exert direct regulation<sup>11</sup>. Instead, two mechanisms control activity. The first is transcriptional, as exemplified by *Drosophila* rhomboid-1: the constitutively active protease is made only when and where needed<sup>3</sup>. This mechanism has historically served as a

beautiful atlas of epidermal growth factor signal initiation during development. The second mechanism is centred on controlling access to substrate by segregating it from protease<sup>11</sup>. Malaria, for example, sequesters adhesins in secretory organelles before invasion, while their secretion onto the surface leads to the first encounter with an active rhomboid protease<sup>7</sup>.

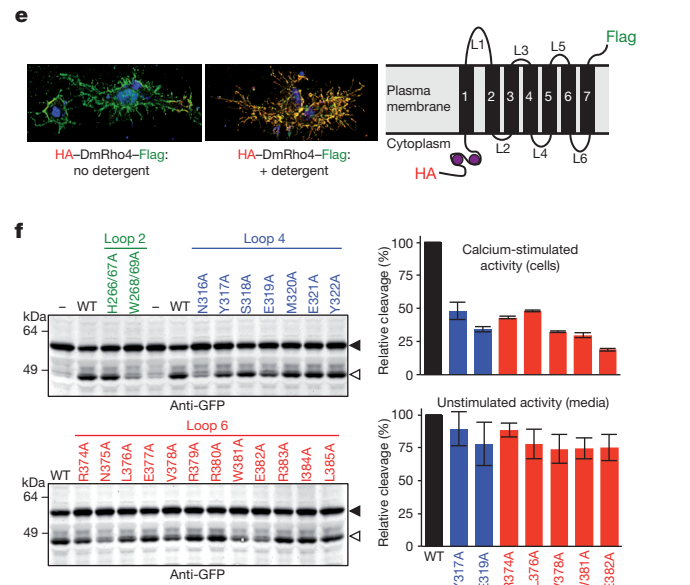
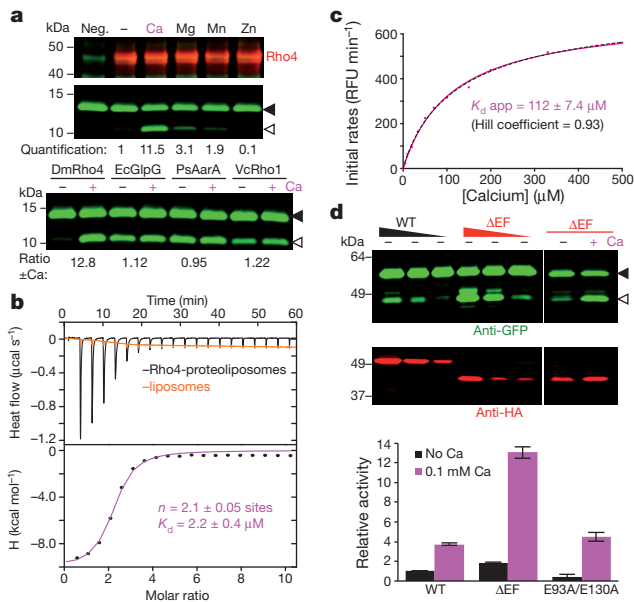
The key property missing from these two mechanisms is the ability to respond rapidly to changing conditions: transcriptional and cell localization changes are ill adapted to provide immediate responses that are hallmarks of cell signalling. Moreover, it is essentially unprecedented for proteases to be devoid of direct enzymatic regulation in the cell, raising the possibility that this apparent discrepancy reflects our lack of understanding rather than absence of a regulatory mechanism.

Although *Escherichia coli* rhomboid protease GlpG has served as a tractable model for studying the structure–function of intramembrane proteolysis<sup>12</sup>, no information is available on its cellular role. This knowledge gap prohibits deciphering regulatory mechanisms. Instead, as a new approach to this question, we searched for rhomboid



**Figure 1 | Calcium rapidly stimulates intramembrane proteolysis in *Drosophila* cells by endogenous DmRho4.** **a**, Diagram comparing the predicted calcium-binding loop residues of DmRho4 to an EF-hand consensus (in red). **b**, Calcium ionophore treatment of *Drosophila* S2R<sup>+</sup> cells induced cleavage of GFP–Spitz, but not its cleavage-site mutant, by endogenous DmRho4. Graph shows expression levels of *Drosophila* rhomboid genes in S2R<sup>+</sup> cells (RNA-seq data from modENCODE, <http://modencode.org>). AU, arbitrary units. **c**, Ionophore-induced Spitz cleavage was detectable within 5 min (red triangle) and linear for 3 h. **d**, RNAi knockdown of DmRho4 but not of DmRho1 abrogated calcium-induced cleavage of GFP–Spitz. **e**, Plasmid expression of DmRho4 rescued calcium-induced cleavage of GFP–Spitz in S2R<sup>+</sup> cells undergoing RNAi. **f**, Calcium-stimulated Spitz cleavage abolished by DmRho4 RNAi could not be rescued by DmRho1 overexpression. All images are anti-GFP western analyses, with substrate and cleavage bands denoted by black or open triangles, respectively, and non-specific bands marked by 'x' (see Fig. 3d for untransfected cells).

<sup>1</sup>Howard Hughes Medical Institute, Department of Molecular Biology & Genetics, Johns Hopkins University School of Medicine, Room 507 PCTB, 725 North Wolfe Street, Baltimore, Maryland 21205, USA.



**Figure 2 | Calcium directly regulates intramembrane proteolytic activity of DmRho4.** **a**, Proteolysis assay with pure reconstituted DmRho4 with or without a panel of 1 mM divalent metal ions (upper panels), and four different rhomboid enzymes reconstituted into proteoliposomes with or without calcium (lower panel). EcGlpG is from *E. coli*, PsAarA is from *Providencia stuartii*, and VcRho1 is from *Vibrio cholerae*. **b**, Analysis of calcium binding to DmRho4 in proteoliposomes by isothermal titration calorimetry (upper graph shows the thermograms, lower graph is the liposome-subtracted quantification). **c**, Calcium titration analysis of DmRho4 proteolysis using an inducible real-time reconstitution assay<sup>16</sup>. Black dashed line shows an alternative fit with an optimal Hill coefficient. RFU, relative fluorescence units. **d**, Titration of wild-type (WT) and  $\Delta$ EF DmRho4 in S2R<sup>+</sup> cells comparing basal, unstimulated Spitz cleavage (left panel), and Spitz cleavage by DmRho4- $\Delta$ EF with or without calcium ionophore (right panel; see also Extended Data Fig. 2a). Lower graph shows *in vitro* activity of wild type and EF-hand mutants

of DmRho4 in proteoliposomes (error bars, s.d. for experimental replicates). **e**, Topology of 3 $\times$ HA-DmRho4-Flag in S2R<sup>+</sup> cells as assessed by deconvolution immunofluorescence. The N-terminal HA-tag (red) was inaccessible while the C-terminal Flag tag (green) was accessible in the absence of detergent, indicating that the N terminus is cytosolic while the C terminus of DmRho4 is extracellular (blue marks nuclei). **f**, Ability of DmRho4 cytosolic loop mutants to cleave GFP-Spitz in response to calcium ionophore stimulation in S2R<sup>+</sup> cells was quantified by anti-GFP western analysis (see also Extended Data Fig. 2b for DmRho4 levels). Graphs show activity of selectively compromised loop 4 and 6 mutants under calcium-stimulated conditions in cells (upper graph) versus unstimulated conditions (lower graph, measured as cleavage product accumulation in culture media after 24 h; see also Extended Data Fig. 2c). Error bars, s.d. for experimental replicates. Filled and open triangles denote substrate and cleavage bands, respectively, throughout.

proteins that contain additional domains with precedent for regulating protein activity and focused on a conserved subset of over two dozen animal rhomboid enzymes with EF-hand domains appended to their cytosolic amino (N) termini (Fig. 1a and Extended Data Fig. 1). EF-hands are helix-loop-helix motifs in which calcium binding at the loop serves either a structural or a regulatory role. In the latter, calcium binding separates the helices and exposes a new surface for binding a regulatory partner<sup>13</sup>. EF-hands typically occur in pairs to form a stable helical bundle, perhaps the best characterized of which are the EF-hands of calmodulin<sup>13</sup>.

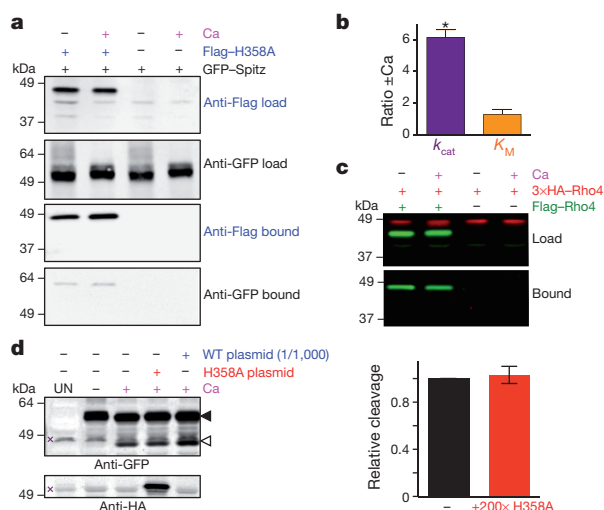
Since rhomboid function is best understood in *Drosophila*, we sought to study the EF-hand containing *Drosophila* rhomboid-4 (DmRho4) under physiological conditions by searching for cell lines that endogenously express DmRho4. We focused on the well-characterized S2R<sup>+</sup> cell line, which also expresses the housekeeping mitochondrial rhomboid and low amounts of DmRho1, but no other rhomboid (Fig. 1b). Treating S2R<sup>+</sup> cells with a calcium ionophore potently stimulated processing of the epidermal growth factor ligand Spitz by more than 50-fold, but not its transmembrane mutant (Fig. 1b), and processing was rapid, becoming detectable within 5 min (Fig. 1c). Targeting DmRho4 with RNA interference (RNAi) removed this stimulation completely, while parallel RNAi against DmRho1 had no effect whatsoever (Fig. 1d). Importantly, exogenous expression of DmRho4 lacking the RNAi target sequence fully rescued calcium stimulation (Fig. 1e). Finally, calcium activated DmRho4 proteolysis directly and not other steps such as Spitz trafficking, because even high levels of DmRho1 could not substitute for removing DmRho4 during calcium stimulation (Fig. 1f). Calcium therefore

triggers potent and rapid proteolysis by DmRho4 under physiological conditions.

Our goal was to study direct regulation of rhomboid enzymes. We therefore next tested the unlikely possibility that calcium directly regulates pure DmRho4 expressed and purified from bacteria and reconstituted into liposomes. Remarkably, addition of calcium directly stimulated intramembrane proteolysis of DmRho4 by more than tenfold, but not a panel of other rhomboid proteases, without any other protein factors present (Fig. 2a). This stimulation was selective since other divalent metal ions failed to have this effect.

We used isothermal titration calorimetry to characterize the thermodynamic basis of calcium binding to the EF-hand domain (Fig. 2b). The resulting thermograms revealed two sites for calcium binding with similar dissociation constants,  $K_d$ , of  $2.1 \pm 0.05 \mu\text{M}$ . Interestingly, a calcium titration revealed stimulation of DmRho4 protease activity (Fig. 2c) could be fitted with a binding isotherm with an apparent  $K_d$  of  $112 \pm 7.4 \mu\text{M}$ , implying that calcium binding to at least a third, lower affinity site (beyond detection by isothermal titration calorimetry) is required for enzyme activation. To assess this further, we deleted the EF domain entirely ( $\Delta$ EF), and found DmRho4 became dysregulated in cells, with its basal activity elevated approximately tenfold (Fig. 2d). This effect was direct because pure DmRho4 $\Delta$ EF also displayed elevated activity *in vitro*. However, calcium still stimulated proteolysis of the E93A+E130A mutant that abrogates binding of calcium to EF-hands, DmRho4 $\Delta$ EF, and DmRho4 $\Delta$ N missing its entire N terminus (Fig. 2d and Extended Data Fig. 2a). These observations further indicate that another calcium-binding site outside its N-terminal domain is the basis of DmRho4 stimulation, while the EF-hand domain functions to limit activation.





**Figure 3 | Intermolecular interactions do not mediate DmRho4 regulation by calcium.** **a**, Anti-Flag co-immunoprecipitation analysis of catalytically inactive Flag-DmRho4-H358A and GFP-Spitz from S2R<sup>+</sup> cells untreated or treated with calcium ionophore. **b**, Effect of calcium on the steady-state kinetic parameters of intramembrane proteolysis by reconstituted DmRho4 (mean  $\pm$  s.d. of three independent experiments, compared using a paired *t*-test). **c**, Anti-Flag co-immunoprecipitation of Flag-DmRho4 and HA-DmRho4 co-expressed in S2R<sup>+</sup> cells untreated or treated with calcium ionophore. **d**, Overexpressing catalytically inactive DmRho4 (H358A) had no effect on the calcium-stimulated activity of endogenous DmRho4 in S2R<sup>+</sup> cells (compare cleavage bands, denoted by open triangle, in lanes 3 versus 4, and quantified in graph on right). Expressing low levels of wild-type 3 $\times$ HA-DmRho4 (1/1,000 amount of input plasmid) was used to quantify the level of endogenous DmRho4 (by comparing protease activity) relative to 3 $\times$ HA-DmRho4-H358A expression (by comparing anti-HA signals). UN indicates S2R<sup>+</sup> cells not transfected with GFP-Spitz ('x' denotes non-specific bands).

To begin mapping the calcium-binding site(s) responsible for stimulating proteolysis, we first examined the topology of DmRho4 in S2R<sup>+</sup> cells (Fig. 2e). An antibody accessibility approach revealed the N terminus of DmRho4 resides in the cytosol while the carboxy (C) terminus is extracellular, indicating that loops 2, 4, and 6 are cytosolic and thus candidates for calcium binding. We mutated all 24 residues in these three cytosolic loops to alanine and assessed their activity in *Drosophila* cells (Fig. 2f and Extended Data Fig. 2b, c). The mutants fell into three classes: most did not affect activity, three perturbed both stimulated and unstimulated activity and were therefore probably structural mutants, and seven selectively compromised calcium-stimulated proteolysis but not DmRho4 structure (Extended Data Fig. 2d), as revealed in a sensitive thermostability assay<sup>14</sup>. The latter localized to loops 4 and 6, indicating that calcium binding at a site formed by loops 4 and 6 specifically activates the intramembrane enzyme core, but how?

We used complementary biochemical and enzymatic approaches to decipher the mechanism of calcium stimulation. Although calcium binding could expose an exosite for substrate binding<sup>15</sup>, we did not detect any increased interaction between DmRho4 and substrate in the presence of calcium by co-immunoprecipitation analysis either from *Drosophila* cells (Fig. 3a) or from proteoliposomes (Extended Data Fig. 3a). This was also functionally corroborated by kinetic analysis using an inducible real-time reconstitution assay for studying catalysis directly within the membrane<sup>16</sup>; calcium stimulated the catalytic rate of DmRho4 at least sixfold, but did not increase affinity of protease for substrate (Fig. 3b).

Recently rhomboid proteases have been postulated to exist as dimers<sup>17</sup>. In contrast to this model, co-expressing two DmRho4 molecules carrying different epitope tags under calcium signalling conditions in *Drosophila* cells (Fig. 3c) or co-reconstituted into

proteoliposomes (Extended Data Fig. 3b) did not result in dimerization. Functionally, we also did not observe calcium stimulation of protease activity *in trans*, which is a classical test of allostery resulting from oligomerization: a catalytically inactive rhomboid enzyme that can still bind calcium could not stimulate the activity of a DmRho4 enzyme carrying a mutation that compromised calcium binding (Extended Data Fig. 3c).

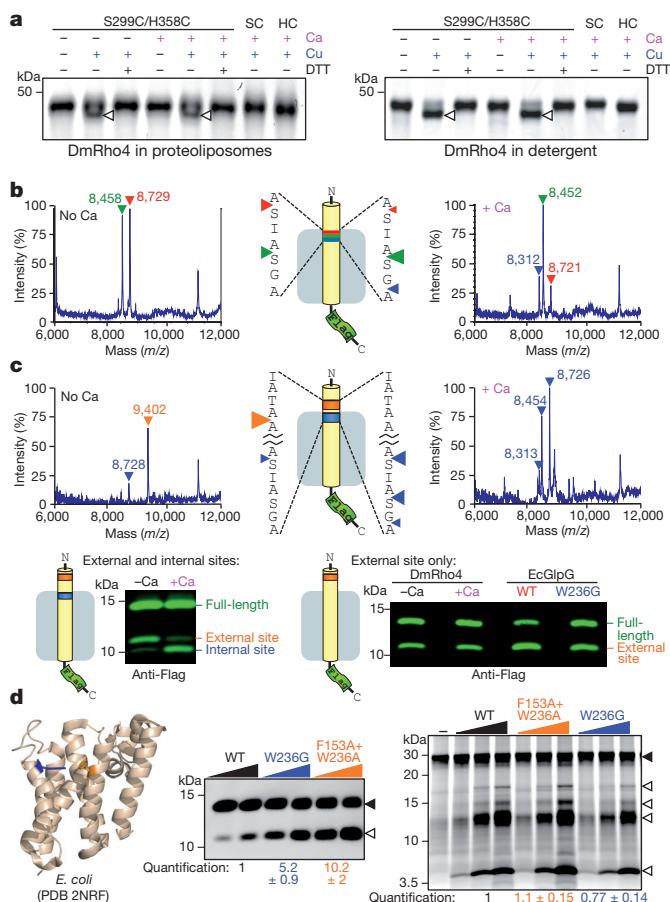
These observations collectively suggest that DmRho4 is not regulated by interaction with any other proteins. To test this further in *Drosophila* cells, we overexpressed inactive DmRho4 to outcompete the endogenous enzyme for any binding partners (Fig. 3d). Remarkably, about 200-fold more inactive DmRho4 had no effect whatsoever on the ability of the endogenous DmRho4 to process Spitz. This observation, in particular, indicates that rhomboid regulation under physiological conditions is not mediated by dimerization, substrate affinity, or additional factors (although factors that fine tune responses in different contexts remain possible).

In contrast to intermolecular target binding, calcium could directly stimulate the activity of a single DmRho4 enzyme through intramolecular allostery. Calpains are the precedent for this type of activation, with calcium binding resulting in a conformation change that aligns the catalytic residues<sup>18</sup>. Since the structure of a eukaryotic rhomboid enzyme has never been solved, we used a biochemical cross-linking approach to test whether calcium aligns the catalytic residues of DmRho4. Cysteines installed at the catalytic serine and histidine positions had no effect on the structural stability of DmRho4 (Extended Data Fig. 4a) and could readily and reversibly be oxidized to form a disulphide bridge (Fig. 4a). Importantly, calcium did not affect the amount of cross-linking, revealing that the DmRho4 catalytic residues are pre-aligned with no influence from calcium binding.

Stimulation of the catalytic rate constant  $k_{cat}$  by calcium was strikingly reminiscent of the increase in  $k_{cat}$  measured for gate-open mutants of GlpG<sup>16</sup>. This is an attractive parallel, because gate-opening is the rate-limiting step for rhomboid intramembrane proteolysis, and the loops that we predict bind calcium also connect the presumed transmembrane gating helix to the rest of the enzyme. One consequence of gate-opening is that substrates can enter deeper into the protease active site, which is reflected in a shift of cleavage site deeper into the substrate transmembrane segment<sup>19</sup>. Accordingly, calcium-stimulated proteolysis shifted the cleavage site 3 residues deeper into the transmembrane segment for DmRho4 but not other rhomboid proteases, consistent with calcium specifically stimulating DmRho4 proteolysis by facilitating gate-opening (Fig. 4b).

To explore the functional consequence of this shift, we examined proteolysis of a series of transmembrane substrates that we engineered to have cleavage sites inside the membrane, outside the membrane, or both in the same molecule. Remarkably, the external site was used very well in the absence of calcium, while the intramembrane site was barely cleaved (Fig. 4c and Extended Data Fig. 4b). Moreover, loop 4 and 6 mutants that compromised calcium regulation in cells also readily cleaved the external site (Extended Data Fig. 4c), which independently confirms that the catalytic residues are competent for catalysis in the absence of calcium. However, addition of calcium shifted the cleavage site from the external site almost exclusively to the intramembrane site for DmRho4 while having no effect on the cleavage site selection of other rhomboid enzymes (Fig. 4c and Extended Data Fig. 4d).

Although no information is available on how *E. coli* GlpG is regulated, we also extended these analyses to this widely studied enzyme by comparing wild type with gate-open mutants<sup>20</sup>. Both wild-type and gate-open GlpG cleaved the external site with similar efficiency (Fig. 4c), while proteolysis at the intramembrane site was specifically stimulated by gate-open mutants (Fig. 4d). Moreover, although gate-open mutants of *E. coli* GlpG stimulated proteolysis of transmembrane substrates by approximately tenfold (Fig. 4d), a soluble casein substrate that approaches the active site from above (not laterally from the membrane) was not proteolysed at any higher level (Fig. 4d).



**Figure 4 | Regulation of intramembrane proteolysis by lateral substrate gating.** **a**, Calcium did not enhance disulphide crosslinking of cysteines installed at the catalytic serine and histidine positions of DmRho4 (triangles indicate crosslinked products). **b**, Shift in substrate cleavage site generated by reconstituted DmRho4 with or without calcium as analysed by mass spectrometry following anti-Flag immunoaffinity capture. Masses and triangles indicating cleavage sites are colour-matched. **c**, Processing of a substrate carrying intramembrane (blue) and external (orange) cleavage sites in APP-Flag that was co-reconstituted into liposomes with DmRho4, and assayed with or without calcium. Reactions were analysed by mass spectrometry (top) and quantitative western blotting (lower panels). Processing of a substrate carrying only the external cleavage site was compared for DmRho4 with or without calcium, and for EcGlpG versus its gate-open mutant W236G. **d**, Quantitative cleavage analysis of wild type and gate-open mutants of EcGlpG for the APP-Spi7-Flag transmembrane substrate and a soluble BODIPY-casein substrate. Full-length substrate (filled triangles) and cleavage products (open triangles) are indicated.

Therefore, cleavage outside the membrane by rhomboid proteases is readily possible, but this type of cleavage is difficult to regulate. A similar analysis of DmRho4 will require identifying gate-open mutants, which should no longer respond to calcium. Lack of a DmRho4 structure and <12% sequence identity with GlpG has hindered this approach.

In summary, we discovered that calcium binding directly, rapidly, and potentially stimulates DmRho4 intramembrane proteolysis by facilitating substrate access to the internal active site (i.e. gating). In addition to revealing, to our knowledge, the first mechanism for directly regulating any intramembrane protease, these observations resolve the long-standing mystery of why these proteases gate substrate entry. Gating has been controversial, because rhomboid proteases can be made to process substrates without lateral gate-opening<sup>21</sup>, and cleavage can be moved to external sites in substrates<sup>22</sup>. In doing so, it was questioned why intramembrane gate-opening occurs, if at all. Our observations now reveal that intramembrane substrate gating is a

means of enzyme regulation, not a necessary step in catalysis: cleavage outside the membrane is possible, but cannot be tightly regulated directly.

The ability of cytosolic regions to regulate substrate gating at a distant, intramembrane site also suggests deeper organizational features within rhomboid architecture that are only beginning to be studied<sup>23</sup>. Interestingly, the 'non-canonical' mode of calcium binding to DmRho4 loops is reminiscent of synaptotagmin activation, which also involves loops that are placed into close contact with lipid molecules. In fact, rhomboid activation may also involve lipids, which might explain the apparent high calcium levels needed for full enzyme activation *in vitro*: synaptotagmin exhibits an intrinsic  $K_d$  of 530  $\mu$ M for calcium at the C2 site that decreases to 3–4  $\mu$ M when appropriate lipids are present<sup>24</sup>. Although ultimately structural analysis is required to reveal the precise architecture of calcium binding, likely involvement of lipid, and impact on gating in DmRho4, so far no rhomboid enzyme with an intact extramembraneous domain has produced well-diffracting crystals. Our studies provide incentive to move beyond GlpG and focus structural biology efforts on these more complex rhomboid proteases.

A particularly exciting implication of these enzymatic properties is that rhomboid proteases can directly integrate upstream signals from other signalling pathways. In fact, this may have medical implications, since Ventrhold/RHBDL3, a human rhomboid that contains potential calcium-binding residues in its EF-hands and cytoplasmic loops, is expressed in the nervous system<sup>25</sup> and may be linked to a mental retardation syndrome<sup>26</sup>. In this light, studying upstream regulation provides a powerful new approach towards revealing biological functions of rhomboid proteases that have evaded discovery. In fact, previous efforts could have missed important roles because they were studying rhomboid functions under unstimulated conditions. Activation is not limited to calcium signalling, since a diversity of recognizable domains have been appended to different rhomboid proteins including zinc fingers,  $\beta$ -propellers, and tetracoordinate repeats<sup>27</sup>. It should be noted that not all extramembraneous domains necessarily serve direct regulatory functions. For example, trafficking signals have been found in the cytosolic domains of parasitic rhomboid enzymes<sup>28</sup>.

Finally, while we focused our studies on rhomboid proteases, substrate gating has been proposed for other intramembrane proteases<sup>29,30</sup>, raising the possibility that gating may be a general mechanism for directly regulating intramembrane proteolysis.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 27 May 2014; accepted 2 March 2015.**

**Published online 11 May 2015.**

1. Brown, M. S., Ye, J., Rawson, R. B. & Goldstein, J. L. Regulated intramembrane proteolysis: a control mechanism conserved from bacteria to humans. *Cell* **100**, 391–398 (2000).
2. De Strooper, B. *et al.* A presenilin-1-dependent  $\gamma$ -secretase-like protease mediates release of Notch intracellular domain. *Nature* **398**, 518–522 (1999).
3. Bier, E., Jan, L. Y. & Jan, Y. N. *rhomboid*, a gene required for dorsoventral axis establishment and peripheral nervous system development in *Drosophila melanogaster*. *Genes Dev.* **4**, 190–203 (1990).
4. Lichtenthaler, S. F., Haass, C. & Steiner, H. Regulated intramembrane proteolysis—lessons from amyloid precursor protein processing. *J. Neurochem.* **117**, 779–796 (2011).
5. Chavez-Gutierrez, L. *et al.* The mechanism of  $\gamma$ -secretase dysfunction in familial Alzheimer disease. *EMBO J.* **31**, 2261–2274 (2012).
6. Golde, T. E., Koo, E. H., Felsenstein, K. M., Osborne, B. A. & Miele, L. gamma-Secretase inhibitors and modulators. *Biochim. Biophys. Acta* **1828**, 2898–2907 (2013).
7. Urban, S. Making the cut: central roles of intramembrane proteolysis in pathogenic microorganisms. *Nature Rev. Microbiol.* **7**, 411–423 (2009).
8. Urban, S., Lee, J. R. & Freeman, M. A family of rhomboid intramembrane proteases activates all membrane-tether EGF ligands in *Drosophila*. *EMBO J.* **21**, 4277–4286 (2002).
9. Baker, R. P., Wijetilaka, R. & Urban, S. Two *Plasmodium* rhomboid proteases preferentially cleave different adhesins implicated in all invasive stages of malaria. *PLoS Pathog.* **2**, 922–932 (2006).

10. Jin, S. M. *et al.* Mitochondrial membrane potential regulates PINK1 import and proteolytic destabilization by PARL. *J. Cell Biol.* **191**, 933–942 (2010).
11. Morohashi, Y. & Tomita, T. Protein trafficking and maturation regulate intramembrane proteolysis. *Biochim. Biophys. Acta* **1828**, 2855–2861 (2013).
12. Brooks, C. L. & Lemieux, M. J. Untangling structure-function relationships in the rhomboid family of intramembrane proteases. *Biochim. Biophys. Acta* **1828**, 2862–2872 (2013).
13. Bhattacharya, S., Bunick, C. G. & Chazin, W. J. Target selectivity in EF-hand calcium binding proteins. *Biochim. Biophys. Acta* **1742**, 69–79 (2004).
14. Baker, R. P. & Urban, S. Architectural and thermodynamic principles underlying intramembrane protease function. *Nature Chem. Biol.* **8**, 759–768 (2012).
15. Fleig, L. *et al.* Ubiquitin-dependent intramembrane rhomboid protease promotes ERAD of membrane proteins. *Mol. Cell* **47**, 558–569 (2012).
16. Dickey, S. W., Baker, R. P., Cho, S. & Urban, S. Proteolysis inside the membrane is a rate-governed reaction not driven by substrate affinity. *Cell* **155**, 1270–1281 (2013).
17. Sampathkumar, P. *et al.* Oligomeric state study of prokaryotic rhomboid proteases. *Biochim. Biophys. Acta* **1818**, 3090–3097 (2012).
18. Moldoveanu, T. *et al.* A  $\text{Ca}^{2+}$  switch aligns the active site of calpain. *Cell* **108**, 649–660 (2002).
19. Moin, S. M. & Urban, S. Membrane immersion allows rhomboid proteases to achieve specificity by reading transmembrane segment dynamics. *eLife* **1**, e00173 (2012).
20. Urban, S. & Baker, R. P. *In vivo* analysis reveals substrate-gating mutants of a rhomboid intramembrane protease display increased activity in living cells. *Biol. Chem.* **389**, 1107–1115 (2008).
21. Xue, Y. & Ha, Y. Large lateral movement of transmembrane helix S5 is not required for substrate access to the active site of rhomboid intramembrane protease. *J. Biol. Chem.* **288**, 16645–16654 (2013).
22. Strisovsky, K., Sharpe, H. J. & Freeman, M. Sequence-specific intramembrane proteolysis: identification of a recognition motif in rhomboid substrates. *Mol. Cell* **36**, 1048–1059 (2009).
23. Bondar, A. N., del Val, C. & White, S. H. Rhomboid protease dynamics and lipid interactions. *Structure* **17**, 395–405 (2009).
24. Fernandez-Chacon, R. *et al.* Synaptotagmin I functions as a calcium regulator of release probability. *Nature* **410**, 41–49 (2001).
25. Jaszai, J. & Brand, M. Cloning and expression of Ventrhoid, a novel vertebrate homologue of the *Drosophila* EGF pathway gene rhomboid. *Mech. Dev.* **113**, 73–77 (2002).
26. Venturin, M. *et al.* Mental retardation and cardiovascular malformations in NF1 microdeleted patients point to candidate genes in 17q11.2. *J. Med. Genet.* **41**, 35–41 (2004).
27. Koonin, E. V. *et al.* The rhomboids: a nearly ubiquitous family of intramembrane serine proteases that probably evolved by multiple ancient horizontal gene transfers. *Genome Biol.* **4**, R19 (2003).
28. Sheiner, L., Dowse, T. J. & Soldati-Favre, D. Identification of trafficking determinants for polytopic rhomboid proteases in *Toxoplasma gondii*. *Traffic* **9**, 665–677 (2008).
29. Feng, L. *et al.* Structure of a site-2 protease family intramembrane metalloprotease. *Science* **318**, 1608–1612 (2007).
30. Li, X. *et al.* Structure of a presenilin family intramembrane aspartate protease. *Nature* **493**, 56–61 (2013).

**Acknowledgements** This work was supported by National Institutes of Health grant 2R01AI066025, the Howard Hughes Medical Institute, and the David and Lucile Packard Foundation. We are grateful to our colleague and friend A. Holland for use of his deconvolution microscope.

**Author Contributions** S.U. and R.B. designed the research. S.U. performed all *Drosophila* cell biology experiments, while R.B. performed all biochemistry experiments. S.U. wrote the manuscript and R.B. made the figures.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.U. ([surban@jhmi.edu](mailto:surban@jhmi.edu)).



## METHODS

**Rhomboid expression constructs.** The *Drosophila* rhomboid-4 open reading frame (ORF) was subcloned from SD06923 (*Drosophila* Genomics Research Center) into pGEX-6P-1 or pET21 for bacterial expression, or into pRmHa3 for expression in *Drosophila* S2R<sup>+</sup> cells. An N-terminal 3×HA-tag was introduced to allow detection of DmRho4 by anti-HA western analysis. Residue substitutions were introduced by site-directed mutagenesis using Pfu Ultra in a Stratagene 96 Gradient Robocycler (Agilent Technologies) and were confirmed by sequencing the entire ORF. The EF-hand domain of DmRho4 (residues 68–154) or its entire N-terminal domain (residues 1–176) were deleted by site-directed mutagenesis to generate the ΔEF and ΔN mutants, respectively.

***Drosophila* cell culture and manipulation.** *Drosophila* S2R<sup>+</sup> cells (purchased from the *Drosophila* Genomics Research Center) were cultured at 25 °C in Schneider's insect medium (Invitrogen) supplemented with 10% fetal bovine serum (FBS, Sigma). For transfection experiments, S2R<sup>+</sup> cells were seeded into six-well plates and transfected with a total of ~2 μg of plasmid DNA (0.5 μg each of pRmHa3-GFP-Spitz and pRmHa3-Star plus 1 μg pBluescript, and 5–25 ng of pRmHa3-3×HA-DmRho4 if applicable) and 5 μl of XtremeGENE HP (Roche). Transfection complexes were formed at room temperature in 100 μl of DMEM for 15 min before being added dropwise to cells. Expression from the metallothionein promoter was induced the following day with 0.5 mM CuSO<sub>4</sub> for ~24 h. For ionophore experiments, cells were then washed with insect saline (10 mM HEPES pH 7.3, 120 mM NaCl, 5 mM KCl, 32 mM sucrose, 8 mM MgCl<sub>2</sub>, 2 mM CaCl<sub>2</sub>), and incubated with 6 μM ionomycin (Sigma) in insect saline for 2–3 h, unless otherwise indicated. Cells were lysed in reducing Laemmli buffer, the lysates resolved on 4–20% tris-glycine SDS polyacrylamide gels, and electrotransferred to nitrocellulose. GFP-Spitz (or its transmembrane mutant SGA→LLL) and 3×HA-DmRho4 were detected by anti-GFP and anti-HA western analysis, respectively, and fluorescence quantified with an Odyssey infrared laser scanner (LiCor Biosciences). The protein level of endogenous DmRho4 in S2R<sup>+</sup> cells was estimated by titrating wild-type 3×HA-DmRho4 by transfection: matching the level of transfected 3×HA-DmRho4 to the endogenous DmRho4 expression level resulted in a doubling of GFP-Spitz processing (Fig. 3d). To detect basal (unstimulated) DmRho4 activity, Schneider's serum-free media was conditioned for 24 h, and the GFP-Spitz that had been released by proteolysis was quantified by anti-GFP western analysis of media fractions.

***Drosophila* cell RNAi.** Templates for *in vitro* transcription were amplified by PCR with primers corresponding to the divergent N-terminal regions of DmRho1 (ORF nucleotides 1–451) and DmRho4 (ORF nucleotides 1–501) and incorporated T7 promoter sequences. RNA was generated using the RiboMAX T7 kit (Promega) according to the manufacturer's instructions, and purified using the RNeasy protocol (Qiagen). Double-stranded RNA (dsRNA) was formed by mixing equal amounts of each strand in annealing buffer (1 mM HEPES pH 7.3, 0.5 mM EDTA) and boiling for 5 min, followed by slow cooling. dsRNA was analysed by agarose gel electrophoresis. S2R<sup>+</sup> cells seeded in six-well plates were washed with serum-free Schneider's media, and ~25 μg of dsRNA was added to each well containing 1 mL of serum-free Schneider's insect media. After a 1–2 h incubation, 3 mL of Schneider's media + 10% FBS was added, and the cells were incubated at 25 °C for 3 days. On the third day, cells were assayed by transfection as described above.

***Drosophila* cell microscopy.** S2R<sup>+</sup> cells were seeded onto glass coverslips, transfected with 3×HA-DmRho4-1×Flag (which was verified to be proteolytically active and calcium regulated), induced with 0.5 mM CuSO<sub>4</sub> for ~24 h, and fixed in 4% formaldehyde in PBS for 20 min. Cells were blocked with 1% bovine serum albumin in the presence or absence of 0.1% TritonX-100, and stained with 1/200 anti-HA and anti-Flag antibodies overnight. The resulting immune complexes were detected with 1/500 secondary antibodies conjugated to Alexa-488 or Alexa-594 (Molecular Probes), mounted in the presence of DAPI, and imaged on a DeltaVision Elite deconvolution microscope (GE Healthcare) using 0.2 μm optical sections with a 1.4 numerical aperture ×100 objective lens.

**Recombinant protein production.** Wild-type or engineered variants of DmRho4 were expressed as N-terminal GST- or His-tag fusions in *E. coli* BL21(DE3) cells. Cultures were grown in M9<sup>+</sup> minimal media containing 100 μg ml<sup>-1</sup> ampicillin at 37 °C, shaking at 250 r.p.m. When cultures reached an absorbance at 600 nm of 0.5, protein expression was induced with the addition of 50 μM IPTG for 16–18 h at 16 °C. Bacterial cell lysates were prepared using a French pressure apparatus operating at 10,000 p.s.i. and cell membranes were pelleted by ultracentrifugation at 350,000g for 30 min. DmRho4 was solubilized from membranes in 2% dodecyl-β-D-maltoside (DDM) for 1 h at 4 °C, followed by ultracentrifugation at 350,000g for 30 min to remove insoluble material. GST-DmRho4 was affinity-purified on glutathione-sepharose and eluted by PreScission protease (GE Healthcare) cleavage of the GST-tag at 4 °C overnight, while His-Rho4 was affinity-purified with Ni-NTA (Qiagen) or His-tag resin (Roche) and eluted using imidazole as

recommended by the manufacturers. The N-terminal His-tag was removed by thrombin cleavage at 4 °C overnight. Bacterial rhomboid proteases were expressed as N-terminal GST fusion proteins and purified by glutathione-sepharose affinity as described above. The APP-Spi7-Flag substrate was produced as described previously<sup>14,31,32</sup>. Briefly, expression of APP-Spi7-Flag was induced for 2–3 h at 37 °C in BL21(DE3) cells with 1 mM IPTG. Cells were lysed using a French pressure apparatus and lysates were cleared of cellular debris by centrifugation at 3,000g for 10 min. The substrate was affinity purified using anti-Flag M2 agarose (Sigma). A variant of APP-Spi7-Flag with an additional rhomboid cleavage site outside the transmembrane domain was constructed by inverse PCR to substitute residues 20–25 (FFAEDV) of APP-Spi7-Flag with the sequence IATAAF from *P. stuartii* TatA. APP-TatA-Flag had only the external TatA site.

**Intramembrane proteolysis assays.** Rhomboid proteases were co-reconstituted with substrates into liposomes using the inducible reconstitution system that we described recently<sup>16</sup>. Bacterial rhomboid proteases were reconstituted in liposomes formed from an *E. coli* polar lipid extract, while DmRho4 was reconstituted in liposomes formed from a yeast polar lipid extract or 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine lipids (Avanti Polar Lipids). All enzymes were assayed for 1 h at 37 °C, except DmRho4, which was assayed at 25 °C for 2–4 h. APP+Spi7-Flag reaction products were resolved by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and quantified by anti-Flag western analysis using an Odyssey infrared laser scanner (LiCor Biosciences), while products for the real-time assay using fluorescein isothiocyanate (FITC)-TatA were quantified using a Synergy H4 Hybrid plate reader (Biotek) scanning once per minute. Proteolysis assays were supplemented with 0.5 mM calcium unless otherwise indicated. For calcium titration experiments, 10 pmol of DmRho4 and 200 pmol FITC-TatA were co-reconstituted into 30 μg yeast liposomes and total calcium was titrated from 0 to 1 mM. For kinetic analysis, 10 pmol of Rho4 was titrated against 15–600 pmol FITC-TatA substrate in the presence or absence of 0.5 mM calcium; initial reaction rates were extracted and fitted to a Michaelis-Menten model using Prism software.

**Casein cleavage assay.** Wild-type GpG and engineered variants were assayed in reactions consisting of 6, 12, or 24 pmol enzyme and 50 μg ml<sup>-1</sup> BODIPY FL casein (Life Technologies) in 50 mM Tris pH 7.4, 150 mM NaCl and 0.1% DDM for 1 h at 37 °C. Reactions were quenched by adding an equal volume of 2× Tricine sample buffer, resolved by SDS-PAGE on 16% Tricine gels (Life Technologies) and imaged using a Typhoon Imager (GE Healthcare) with settings of 488 nm for excitation and 526 nm for emission.

**Isothermal titration calorimetry.** Isothermal titration calorimetry used a Microcal iTC<sub>200</sub> instrument (GE Healthcare). The reaction cell contained 20–30 μM DmRho4 in proteoliposomes and the reference cell contained water. A motorized syringe loaded with 1 mM calcium was used to perform successive 2 μl injections into the reaction cell at 25 °C. Control experiments titrating calcium against liposomes were performed to determine the heat of titrant dilution, which, subtracted from the heat of reaction, yielded the effective heat of calcium binding. Data were fitted using Origin analysis software.

**Thermostability analysis.** Wild-type and engineered variants of DmRho4 were subjected to quantitative thermostability analysis as described previously<sup>14</sup>. Briefly, pure DmRho4 was diluted to 5 μM in 50 mM Tris pH 7.4, 150 mM NaCl, 0.1% dodecyl-β-D-maltoside, 0.5% or 1.2% nonyl-glucoside, heated from 25 °C to 85 °C at a rate of 0.2 °C per minute, and the differential static light scattering was quantified every 0.5 °C in a Stargazer-384 instrument (Harbinger Biotech). Light scattering data were fitted to a two-state Boltzmann curve using Stargazer BioActive software to derive transition temperature midpoints (*T*<sub>m</sub>).

**Rhomboid-substrate co-immunoprecipitation.** S2R<sup>+</sup> cells were co-transfected with catalytically inactive pRmHa3-Flag-DmRho4-H358A and pRmHa3-GFP-Spitz or, as a negative control, with pRmHa3-GFP-Spitz alone, and were either untreated or treated with 6 μM ionomycin in insect saline or Ca-free insect saline, as described above. Cell lysates were solubilized for 1 h at room temperature in 25 mM Tris pH 7.4, 150 mM MgCl<sub>2</sub>, Complete EDTA-free protease inhibitor cocktail (Roche), and 0.25% DDM in either the presence of 1 mM CaCl<sub>2</sub> or 1 mM MgCl<sub>2</sub>. Cell debris was removed by centrifugation at 16,000g for 20 min at 4 °C. Immunoprecipitations were done using anti-Flag M2 agarose for 1 h at room temperature. Beads washed in the presence of 1 mM CaCl<sub>2</sub> or 1 mM MgCl<sub>2</sub> were resuspended in SDS sample buffer, then load and bound fractions were resolved by SDS-PAGE followed by anti-Flag/anti-GFP western analysis. For co-immunoprecipitation analysis in proteoliposomes, catalytically inactive HA-tagged DmRho4 (S299A) was co-reconstituted with APP-Spi7-Flag as described above in buffer consisting of 50 mM Tris pH 7.5, 150 mM NaCl, and either no calcium or 0.5 mM CaCl<sub>2</sub>. Proteoliposomes were incubated for 1 h at room temperature, solubilized with 1% DDM for 30 min at room temperature, and then immunoprecipitation was performed with anti-Flag M2 agarose for 1 h at room temperature in the presence or absence of calcium. Washed beads were

resuspended in SDS sample buffer and the input and bound fractions were compared by anti-HA and anti-Flag western analysis.

**Rhomboid co-immunoprecipitation.** S2R<sup>+</sup> cells were co-transfected with expression constructs encoding triple HA-tagged and Flag-tagged DmRho4, and as a control, with HA-tagged DmRho4 alone as described above, and were then either untreated or treated with 6  $\mu$ M ionomycin in insect saline or Ca-free insect saline. Cells were solubilized with 0.25% DDM (as described above) and immunoprecipitations were performed with anti-Flag M2 agarose for 1 h at room temperature. Beads washed in the presence of 1 mM CaCl<sub>2</sub> or 1 mM MgCl<sub>2</sub> were resuspended in SDS sample buffer. Load and bound fractions were detected by anti-HA and anti-Flag western analysis. Pure samples of HA-tagged and Flag-tagged DmRho4 were also co-reconstituted into proteoliposomes in the absence or presence of 0.5 mM CaCl<sub>2</sub>, then solubilized with 1% DDM for 30 min at room temperature, immunoprecipitated with anti-Flag M2 agarose (as described above) and then subjected to anti-HA/anti-Flag western analysis.

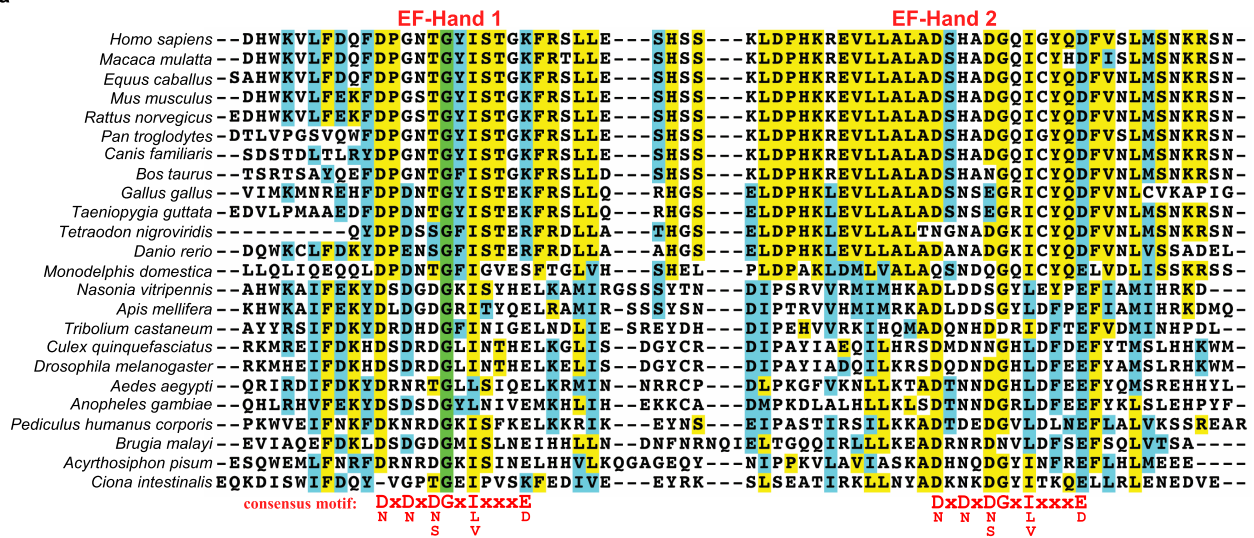
**Disulphide cross-linking.** A cysteine-less mutant of DmRho4 was generated by substituting the three amino-terminal native cysteine residues with serine residues (C104S, C150S, C176S) and the native transmembrane cysteine residue with a valine residue (C334V). Using this construct as a template, the active site residues S299 and H358 were replaced with cysteine residues, either singly, to generate S299C and H358C, or in combination, to generate the double mutant S299C/H358C. Pure cysteine-substituted proteins were reduced by treatment with 5 mM TCEP for 30 min at room temperature and then passed through a Zeba spin

column (Pierce) equilibrated in 50 mM Tris, 150 mM NaCl, 0.1% DDM. Proteins were then oxidized, either in detergent micelles or after reconstitution into proteoliposomes, by the addition of 50 or 100  $\mu$ M copper phenanthroline, respectively, for 15 min at room temperature in the absence or presence of 0.5 mM calcium. Note that owing to the random orientation of rhomboid in reconstituted proteoliposomes, only half of the DmRho4 was accessible to the oxidizing reagent. Control reactions were done in parallel with no copper but with 50 mM DTT to prevent spontaneous oxidation. Reactions were stopped by the addition of SDS-sample buffer. As an additional control, the oxidation observed for the double mutants was reversed by the addition of 50 mM DTT for 10 min to the oxidized proteins in SDS sample buffer. Samples were analysed by SDS-PAGE and stained using either IRDye Blue Protein Stain (LiCor) or Krypton Infrared Protein Stain (Pierce).

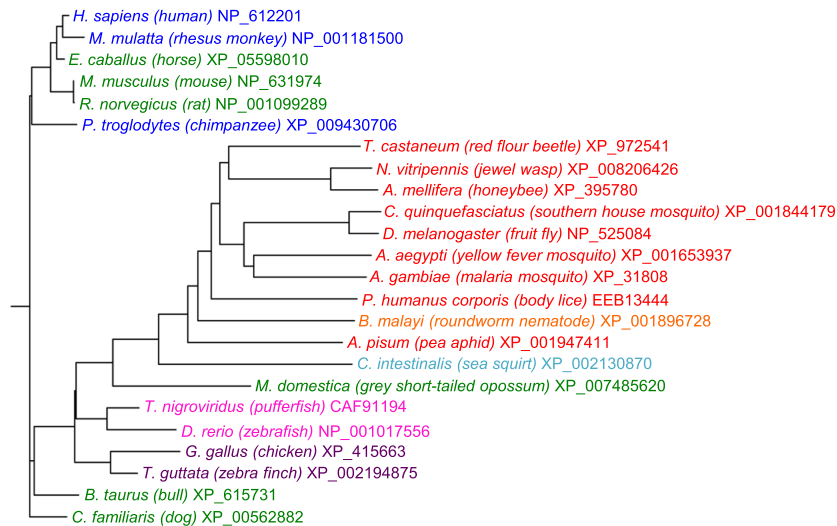
**Mass spectrometry.** Full-length substrate and C-terminal cleavage products were purified from *in vitro* proteolysis assays by anti-Flag immunoaffinity isolation and analysed by matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry using sinapinic acid matrix as described previously<sup>19,32</sup>. No statistical methods were used to predetermine sample size.

31. Urban, S. & Wolfe, M. S. Reconstitution of intramembrane proteolysis *in vitro* reveals that pure rhomboid is sufficient for catalysis and specificity. *Proc. Natl Acad. Sci. USA* **102**, 1883–1888 (2005).
32. Baker, R. P., Young, K., Feng, L., Shi, Y. & Urban, S. Enzymatic analysis of a rhomboid intramembrane protease implicates transmembrane helix 5 as the lateral substrate gate. *Proc. Natl Acad. Sci. USA* **104**, 8257–8262 (2007).

a



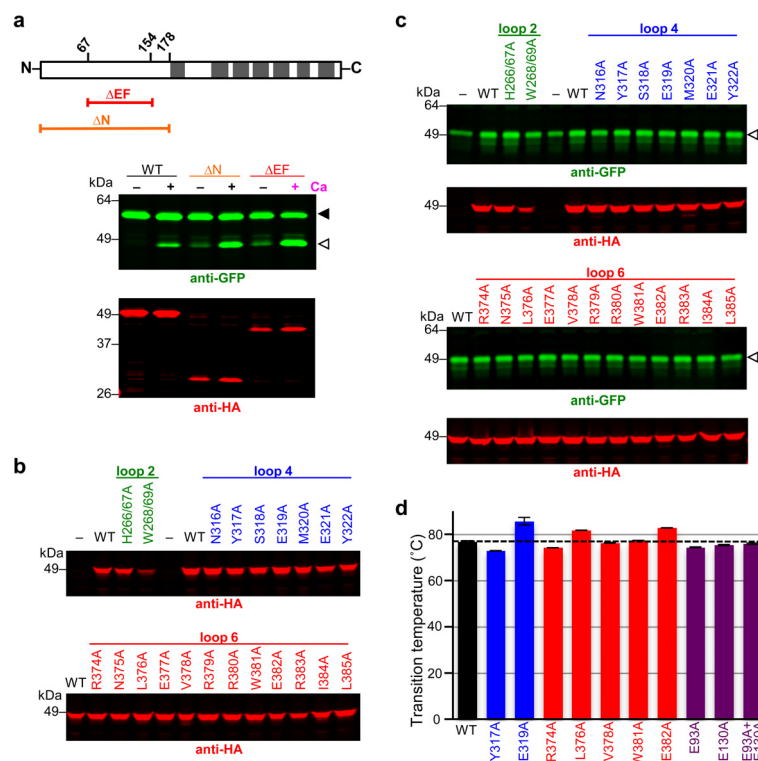
b



**Extended Data Figure 1 | The rhomboid-4 subfamily of rhomboid proteases.** **a**, ClustalW multiple sequence alignment of the conserved N-terminal EF-hand domains of 24 members of the rhomboid-4 subfamily (generated in Biology Workbench, <http://workbench.sdsc.edu>). Identical residues are shaded in green, highly conserved residues in yellow, and similar residues in cyan. The EF-hand calcium-binding loop consensus sequence is

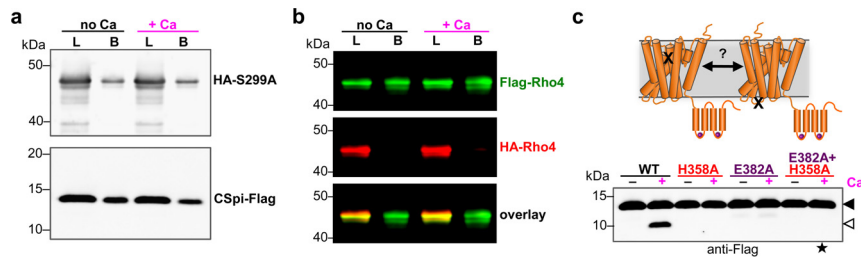
given below the alignment. **b**, Rooted tree of the 24 rhomboid-4 homologues. Genus and species names are colour-coded as follows: primates (blue), other mammals (green), birds (purple), fish (pink), non-vertebrate chordate (cyan), nematode (orange), and insects (red), with vernacular names given in parentheses, followed by National Center for Biotechnology Information (NCBI) accession numbers.





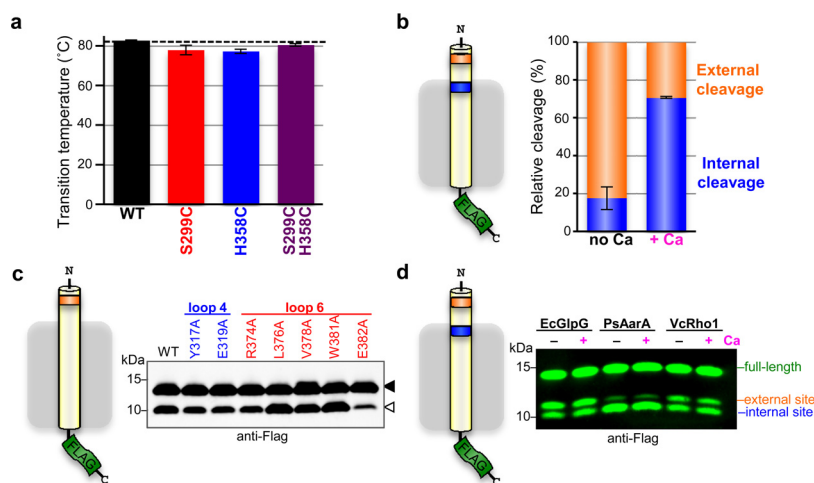
**Extended Data Figure 2 | Activity and thermostability analysis of DmRho4 mutants.** **a**, Comparison of calcium stimulation of DmRho4 versus its EF-hand domain deletion mutant ( $\Delta$ EF), and a mutant lacking the entire cytosolic domain ( $\Delta$ N). Upper diagram shows position of domains (demarcated by residue numbers) and the corresponding deletion constructs. Transmembrane segments are shown as grey rectangles. GFP-Spitz substrate and cleavage products (green bands in the anti-GFP western) are denoted by black and white triangles, respectively. DmRho4 protein levels are shown as red bands (anti-HA western). **b**, Analysis of DmRho4 loop 2, 4, and 6 mutant protein levels from Fig. 2f (calcium stimulation conditions). **c**, DmRho4 loop 2, 4, and 6 mutants

were assayed for cleavage of GFP-Spitz under basal (unstimulated) conditions for ~24 h in the absence of calcium. Cleavage product (green bands, white arrowhead) was detected in media fractions for most of the mutants at levels comparable to the wild-type enzyme. Corresponding DmRho4 protein levels are shown as red bands (anti-HA western analyses). **d**, Wild-type DmRho4 and engineered variants were expressed and purified from bacteria, subjected to quantitative thermal stability analysis, and transition temperature midpoints ( $T_m$ ) were derived (error bars, s.d. of four experimental replicates). The thermal stability of mutant DmRho4 proteases was indistinguishable from that of wild-type DmRho4.



**Extended Data Figure 3 | Calcium does not regulate DmRho4 through intermolecular interactions.** **a**, Anti-Flag co-immunoprecipitation analysis of HA-DmRho4 and APP-Spi7-Flag substrate from proteoliposomes in the presence or absence of 0.5 mM calcium. An inactive mutant of DmRho4 (S299A) was used to facilitate substrate complex isolation. The amount of HA-tagged DmRho4 co-immunoprecipitated with the Flag-tagged substrate was not affected by the presence of 0.5 mM calcium. L, load; B,

bound. **b**, Anti-Flag co-immunoprecipitation of Flag-DmRho4 and HA-DmRho4 from proteoliposomes. HA-tagged DmRho4 failed to co-immunoprecipitate with Flag-tagged DmRho4 in both the absence and presence of 0.5 mM calcium. **c**, Mixing a catalytic mutant (H358A) and a calcium-binding mutant (E382A) cannot rescue calcium stimulation *in trans* (star indicates lane where a product would be expected with the mixed single mutants).



**Extended Data Figure 4 | Lateral substrate gating underlies direct regulation of intramembrane proteolysis.** **a**, Thermostability analysis of single and double cysteine mutants of DmRho4 (error bars, s.d. of four experimental replicates). **b**, Average relative proportions of cleavage at the external cleavage site (orange) compared with the internal cleavage site (blue) are shown for DmRho4 in the absence (no Ca) and presence (+ Ca) of 1 mM calcium (error bars, standard error of replicate experiments). The external site was favoured in the absence of calcium (approximately 80%) while internal cleavage was preferred in the presence of calcium (approximately 70%).

**c**, DmRho4 loop 4 and loop 6 calcium-binding site mutants retained calcium-independent cleavage of a substrate harbouring only an external cleavage site. Full-length substrate (filled triangle) and cleavage product (open triangle) are indicated. **d**, Cleavage of a substrate with external and internal cleavage sites was compared for *E. coli* GlpG, *P. stuartii* AarA, and *V. cholerae* Rho1 in the absence (no Ca) or presence (+ Ca) of 0.5 mM calcium. The relative proportions of cleavage at the two sites varied between the bacterial rhomboid proteases, but in no case did calcium alter the cleavage site preference.



# Structures of actin-like ParM filaments show architecture of plasmid-segregating spindles

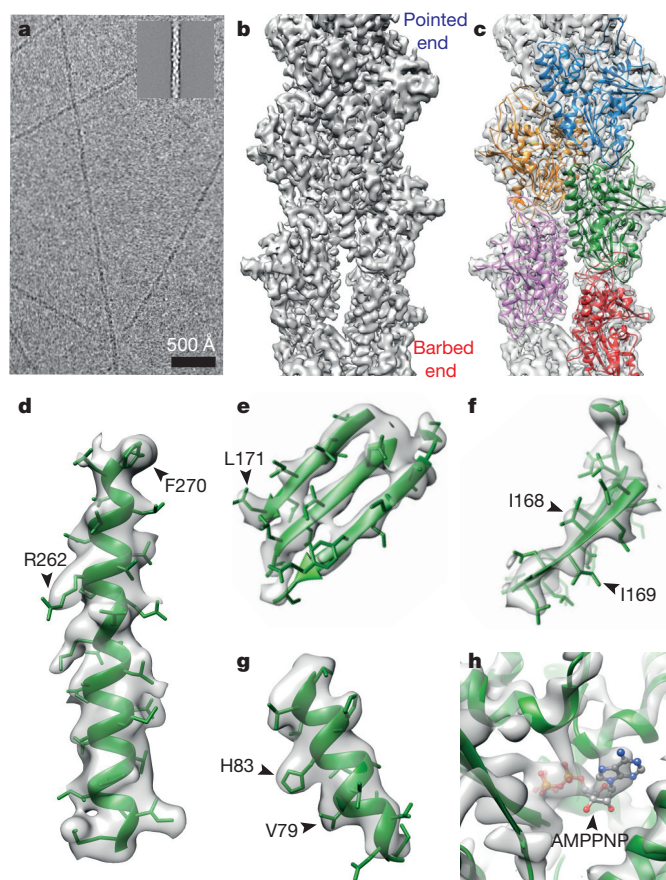
Tanmay A. M. Bharat<sup>1</sup>, Garib N. Murshudov<sup>1</sup>, Carsten Sachse<sup>2</sup> & Jan Löwe<sup>1</sup>

**Active segregation of *Escherichia coli* low-copy-number plasmid R1 involves formation of a bipolar spindle made of left-handed double-helical actin-like ParM filaments<sup>1–6</sup>. ParR links the filaments with centromeric *parC* plasmid DNA, while facilitating the addition of subunits to ParM filaments<sup>3,7–9</sup>. Growing ParMRC spindles push sister plasmids to the cell poles<sup>9,10</sup>. Here, using modern electron cryomicroscopy methods, we investigate the structures and arrangements of ParM filaments *in vitro* and in cells, revealing at near-atomic resolution how subunits and filaments come together to produce the simplest known mitotic machinery. To understand the mechanism of dynamic instability, we determine structures of ParM filaments in different nucleotide states. The structure of filaments bound to the ATP analogue AMPPNP is determined at 4.3 Å resolution and refined. The ParM filament structure shows strong longitudinal interfaces and weaker lateral interactions. Also using electron cryomicroscopy, we reconstruct ParM doublets forming antiparallel spindles. Finally, with whole-cell electron cryotomography, we show that doublets are abundant in bacterial cells containing low-copy-number plasmids with the ParMRC locus, leading to an asynchronous model of R1 plasmid segregation.**

Using electron cryomicroscopic (cryo-EM) images collected on a direct-electron detector, we performed real-space helical reconstruction to elucidate a 4.3 Å structure of ParM filaments assembled with the nucleotide AMPPNP (Fig. 1a–c and Extended Data Fig. 1, Extended Data Table 1 and Supplementary Video 1). Densities corresponding to  $\alpha$ -helices,  $\beta$ -strands and many side chains were clearly observed (Fig. 1d–g). AMPPNP was also observed in our map as strong density, especially on the phosphates (Fig. 1h). No apparent resolution anisotropy was detected in the reconstruction (Extended Data Fig. 1), indicating that the entire ParM protein is rigidly held in the filament. To derive an atomic model of the ParM filament, a previous, monomeric crystal structure of ParM and AMPPNP bound to the tail of ParR (Protein Data Bank (PDB) accession code 4A62) was fitted into the map, and the filament model iteratively rebuilt and all-atom refined using stereochemical restraints with REFMAC.

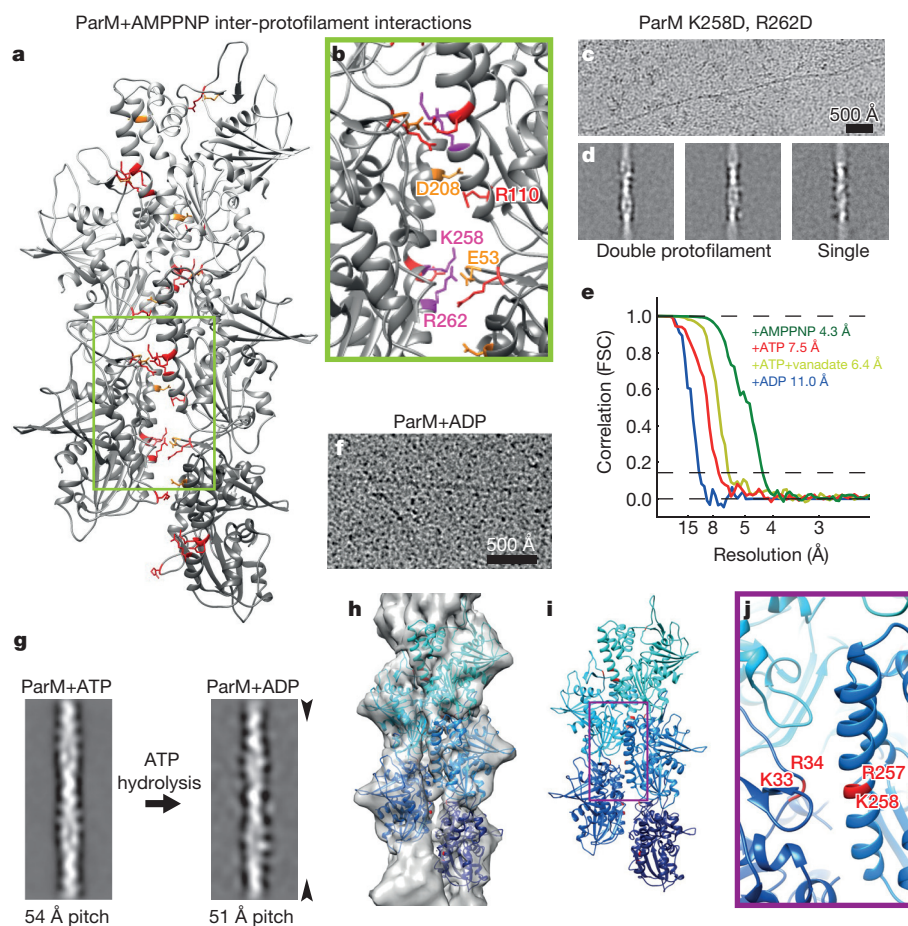
Surprisingly, the two protofilaments (strands) making up the double-helical ParM filament are held together only by salt bridges (Fig. 2a, b, Extended Data Figs 2 and 3 and Extended Data Table 2). The ParM inter-protofilament interface is small (calculated interface area 371 Å<sup>2</sup>) and does not resemble a canonical protein–protein interface containing a hydrophobic core. To demonstrate the validity of this assessment we mutated two positively charged residues within the inter-protofilament interface to aspartic acids (K258D, R262D) and tested what effect this had on the stability of ParM filaments. Filament formation (with AMPPNP) from the resulting mutant protein ParM (K258D, R262D) was inefficient (Extended Data Fig. 3g). The few filaments that were formed were unstable, and tended to be bent (Fig. 2c and Extended Data Fig. 3h). Reference-free class averaging of these filaments showed that even though most of the few observed filaments were double helical like wild-type ParM, some single-helical filaments were also

present (Fig. 2d and Extended Data Fig. 3i). These observations indicate that although the interface between protofilaments in ParM is surprisingly small, it is sufficient for double-helical filament assembly since many identical contacts along the filament contribute to the overall binding energy. Different actin-like proteins show very different filament arrangements, from single (crenactin, possibly<sup>11</sup>) to parallel double helical (left-handed: ParM; right-handed: actin; and non-staggered: MamK<sup>12</sup>) and antiparallel, double straight



**Figure 1 | Cryo-EM reconstruction at 4.3 Å of ParM+AMPPNP filaments.** **a**, Cryo-EM image of ParM+AMPPNP filaments. Inset: class average. This experiment was repeated nine times. **b**, A 4.3 Å reconstruction of the filaments, isosurface contoured at  $2\sigma$  away from the mean (see Extended Data Fig. 1 and Supplementary Video 1). **c**, The same reconstruction as **b**, overlaid with the refined atomic model with individual ParM subunits coloured differently. **d–g**, Enlarged regions of the cryo-EM map showing resolved secondary structure elements and side-chain densities, contoured at  $1\sigma$ . **h**, Density for the nucleotide is stronger than that of the protein (contoured at  $3\sigma$ ).

<sup>1</sup>Structural Studies Division, MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge CB2 0QH, UK. <sup>2</sup>Structural and Computational Biology Unit, European Molecular Biology Laboratory, Meyerhofstrasse 1, Heidelberg 69117, Germany.



**Figure 2 | ParM filaments are made up of two protofilaments held together by salt bridges, which are perturbed when ParM is bound to ADP.** **a**, The refined atomic model of ParM+AMPPNP filaments shows that the protofilaments are held together laterally by salt bridges. Basic residues at the interface are highlighted in red and acidic residues in orange (see Extended Data Table 2). Within the protofilaments' longitudinal interfaces, more extensive hydrophobic interactions are observed (see Extended Data Fig. 2). **b**, A magnified view of **a**. The charges of two basic residues at the interface were inverted by mutation for **c** (K258D, R262D). **c**, The resulting protein formed filaments inefficiently. Cryo-EM image showing filaments of ParM(K258D, R262D) assembled with AMPPNP. This experiment was repeated four times. **d**, In addition to normal double-helical filaments, some single-helical filaments

were observed by image classification and averaging. **e**, Fourier shell correlation (FSC) curves for the four cryo-EM structures presented in this study (see Extended Data Table 1). **f**, Cryo-EM image of ParM+ADP filaments. High protein concentrations were required to obtain these filaments and monomeric proteins can be seen. This experiment was repeated six times. **g**, Comparison of filtered class averages of ParM+ATP and ParM+ADP filaments. Compared with the ATP bound state, the pitch of the ParM+ADP filaments reduced by  $\sim 3$  Å (see Supplementary Video 2). **h**, Cryo-EM reconstruction of ParM+ADP filaments at 11 Å resolution with five copies of the ParM+ADP X-ray structure fitted. **i**, The same pseudo-atomic fit without the cryo-EM density. **j**, A magnified view of the perturbed inter-protofilament interface in the ParM+ADP filaments.

(MreB). We propose that small and simple inter-protofilament contacts could have made it possible to change inter-protofilament arrangements relatively easily during evolution since all these actin-like filaments show similar longitudinal contacts<sup>13</sup>.

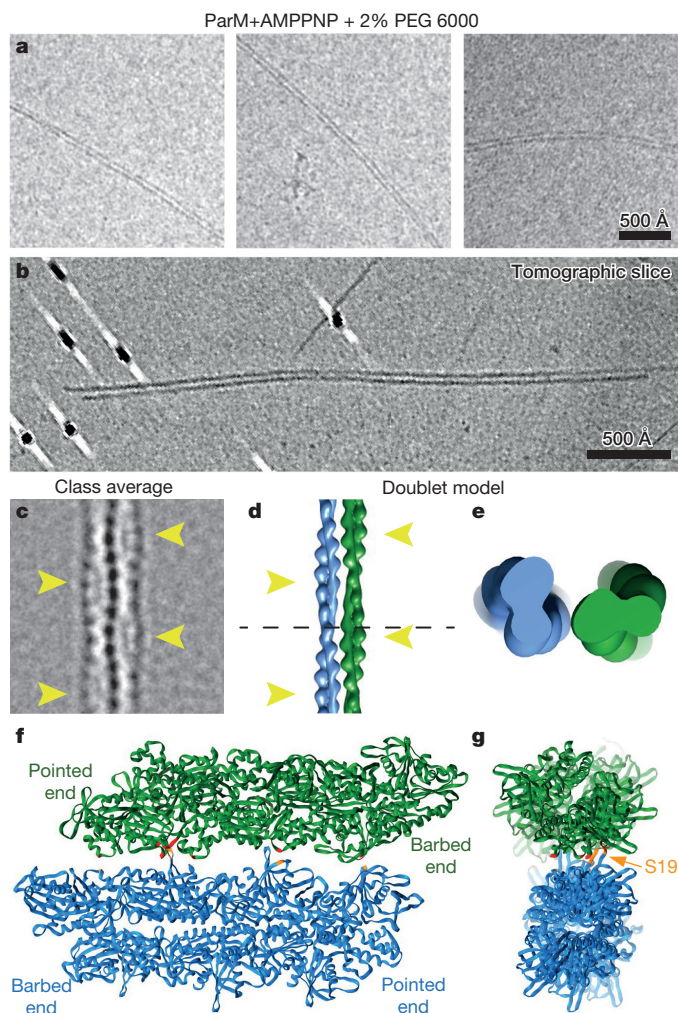
The protofilaments of ParM themselves are held together by an extensive longitudinal contact area ( $\sim 995$  Å<sup>2</sup>), containing both hydrophilic and hydrophobic interactions (Extended Data Fig. 2 and Extended Data Table 2). Actin filaments have also been shown to have the same difference in interface size between the longitudinal and lateral contacts<sup>14–16</sup>. Interestingly, this difference has also been observed in tubulin polymers, microtubules<sup>17</sup>.

The dynamic instability of ParM is caused by intrinsic ATP hydrolysis in the filament and the resulting ADP-bound filament being less stable<sup>18</sup>, while being temporally protected by an ATP cap. We therefore assembled ParM+ATP filaments and obtained a 7.5 Å cryo-EM structure of these filaments (Extended Data Fig. 4). Since the nucleotide state of this structure may be mixed, we devised a way to inhibit the ATPase of ParM with vanadate. Addition of sodium orthovanadate to the ParM+ATP solution retarded filament disassembly and we captured these ParM+ATP+vanadate filaments before complete

disassembly and obtained a 6.4 Å structure (Extended Data Fig. 4). Comparison of the three cryo-EM structures (+AMPPNP, +ATP, +ATP+vanadate) indicates that ParM is held in the same rigid, compact conformation, either until ATP is hydrolysed to ADP or until phosphate is released (Extended Data Fig. 4e, f).

Therefore the state with an expected conformational change should be ADP-bound, and since ParM+ADP has a much higher critical concentration for filament formation, we incubated a concentrated solution of ParM with ADP for cryo-EM (Fig. 2f). This specimen yielded a lower resolution reconstruction at 11 Å. Further refinement was not possible, and adding data did not improve the resolution of the structure (Fig. 2e and Extended Data Table 1). This indicated marked flexibility in the ParM+ADP filaments. Surprisingly, the overall helical pitch of the ParM+ADP filaments is smaller than in the other nucleotide states (ADP: 51 Å versus 54 Å; Fig. 2g and Extended Data Table 1). The previously solved ParM+ADP X-ray structure (PDB 1MWM)<sup>6</sup> was subdivided into its two domains and these were fitted as rigid bodies into the ParM+ADP cryo-EM reconstruction (Fig. 2h, i). Since the helical symmetry of ParM+ADP filaments is different from the ParM+ATP filaments, the interaction of ParM





**Figure 3 | ParM doublets formed *in vitro*.** **a**, Cryo-EM images of ParM doublets formed *in vitro* with crowding agent PEG 6000. This experiment was repeated 15 times. **b**, Slice through an electron cryotomogram (cryo-ET) showing clear lack of super-helicity in the doublets (see Supplementary Video 3). **c**, A two-dimensional class average of the ParM doublet. The thickest parts of double helical ParM filaments have been indicated with yellow arrowheads (see Extended Data Fig. 5). **d**, Model of the doublet, shown in the same orientation as the class average in **c** (see Supplementary Video 4). **e**, An orthogonal, magnified view of the doublet cut at the plane shown as a dashed line in **d**. **f**, Atomic model of the doublet. Residues shown in red in one ParM filament interact with residues in orange in the other filament (see Extended Data Table 2). **g**, An orthogonal view of the doublet, with the filament axes going into the plane of the paper. One of the residues (S19) that forms the doublet interface has been highlighted (see Extended Data Fig. 6).

subunits with each other is also different in the two states. In ParM+ADP filaments, salt bridges at the inter-protofilament interface can no longer be formed and rather repulsing charges are brought close together (Fig. 2j). Additionally, change in helical pitch of the filament may also come with a substantial change in the longitudinal interface. These two factors together could explain why ParM+ADP filaments are less stable, and indicate why ParM filaments rapidly dissociate into monomeric form upon ATP hydrolysis, leading to dynamic instability (Supplementary Video 2).

Having described the structure of the ParM filaments, we then wished to put the structural data in context of the bipolar spindles that segregate plasmid DNA in cells. For bipolar spindles to form, filamentous ParM subunits must engage in another interaction, inter-filament contacts, formed between double-helical filaments. It is known that incubation of ParM filaments with a crowding agent causes them to

bundle<sup>19</sup>. However, bundles are not amenable to high-resolution cryo-EM analysis because of their heterogeneity<sup>20</sup>. To obtain a more defined sample, we titrated ParM+AMPPNP with varying amounts of crowding agent. When 2% polyethylene glycol (PEG) 6000 was added to ParM+AMPPNP, we found that ParM filaments dimerized to form ‘doublets’, containing two double-helical filaments (Fig. 3a and Extended Data Fig. 5a, b). In raw cryo-EM images, doublets appeared as two roughly parallel lines, with no evidence of supercoiling or twisting. Electron cryotomography (cryo-ET) of the doublet specimen confirmed that the filaments do not twist around each other (Fig. 3b and Supplementary Video 3).

We then performed reference-free two-dimensional classification of doublet images (Fig. 3c and Extended Data Fig. 5c). The two ParM filaments in the doublet were perfectly out of phase with each other. When viewed as a projection (in a cryo-EM class average), the thickest part of one filament in the doublet perfectly aligns with the thinnest part of the other double helical filament. We picked small segments along single ParM filaments that formed the doublets and aligned the segments to re-projections of the high-resolution ParM+AMPPNP structure we solved above. Using this alignment, directionality could be assigned to each filament in the doublet. We found that, in 84% of the cases, ParM *in vitro* doublets appeared to be made of two anti-parallel filaments (Extended Data Fig. 5d) while opposite matches were probably due to incorrect assignment of the short segments.

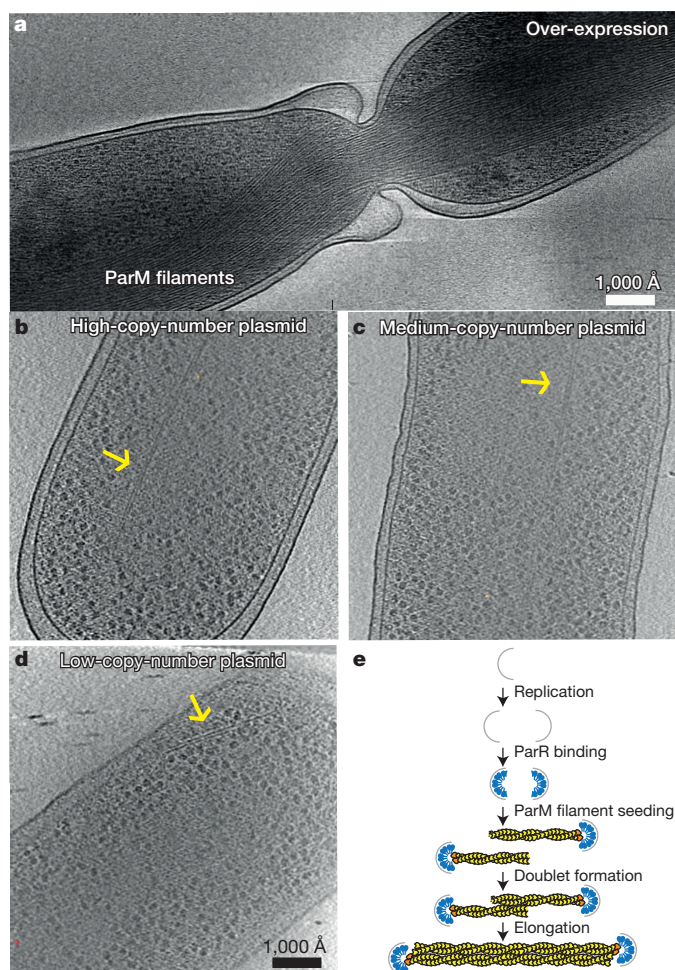
Using the class averages and the directionality assignment, we obtained an averaged model for the ParM doublet (Fig. 3d–g, Extended Data Fig. 5e, f and Supplementary Video 4). Two ParM monomers from adjoining filaments in the doublet model were found to be in a similar orientation as observed in a previous crystal structure of ParM (Extended Data Fig. 6a, b)<sup>3</sup>. The model of the doublet predicts residues in ParM that should be important in doublet formation (Fig. 3f, g and Extended Data Table 2) and confirmed earlier work, including mutations that modulate the strength of the inter-filament contact. One such set of mutations consisted of S19R and G21R<sup>3</sup>. These mutations had been selected previously because they are located the furthest away from the filament axis, essentially sticking out, but are shown here directly to be involved in the inter-filament contact. In line with this, mutant ParM(S19R, G21R) spontaneously formed doublets and bundles (Extended Data Fig. 6c), without any crowding agent present in solution, validating both the previous total internal reflection fluorescence data<sup>3</sup> as well as the current atomic model of the ParM doublet.

Previous imaging by total internal reflection fluorescence microscopy of the reconstituted ParMRC spindles<sup>3</sup> as well as the model of the ParM doublet derived here are *in vitro* experiments. To test whether the doublets have physiological relevance, we visualized ParM filaments inside growing *E. coli* cells. Previously, direct observation of ParM filaments by cryo-EM was only possible by cryo-sectioning of frozen bacterial cells since whole cells were deemed too thick<sup>19</sup>. Importantly, in vitreous sections filaments could only be visualized end-on, not revealing much about the inter-filament contacts. Using new direct electron detectors, signal-to-noise has been dramatically improved, so we aimed at imaging bipolar spindles directly inside cells using whole-cell cryo-ET.

As a first test, we overexpressed a mutant of ParM (D170A) that hydrolysed ATP much more slowly in thin *E. coli* cells. As observed previously in vitreous sections<sup>19</sup>, cryo-ET of these cells (Fig. 4a) allowed unambiguous identification of the overexpressed ParM mutant protein through its tendency to form extremely large bundles.

We then used plasmids with different copy numbers<sup>21</sup>, all of which contained the entire ParMRC locus and transformed them in turn into *E. coli* cells. Cryo-ET of these cells revealed the presence of doublets in all cases (Fig. 4b–d, Supplementary Videos 5 and 6, Extended Data Fig. 7 and Extended Data Table 3). All doublets were roughly aligned with the long cell axis, and were never observed perpendicular to the cell axis. Although bundles were observed in





**Figure 4 | ParM doublets in *E. coli* cells, imaged by cryo-ET.** **a**, A mutant of ParM that hydrolyses ATP more slowly (D170A) was overexpressed in *E. coli* cells. Tomographic slices show large bundles of ParM blocking cell division. This experiment was performed two times. **b**, The ParMRC operon driven from high-copy-number plasmid pDD19. Tomographic slice showing an example of observed doublets. **c**, Tomographic slice for a medium-copy-number plasmid (pKG321). **d**, Tomographic slice for a low-copy-number plasmid, emulating the native low-copy-number R1 plasmids (pKG491, 'mini-R1' replicon) in *E. coli* (see Supplementary Videos 5 and 6 to view entire tomograms). Each experiment with different copy-number plasmids was performed once. **e**, Schematic depicting proposed asynchronous plasmid DNA segregation. Bipolar ParM spindles are seeded when replication has produced two *parC* centromeric regions, still in close proximity. Each seeds one unipolar ParM filament, which then come together in an antiparallel fashion to form the segregating bipolar spindle. Non-productive unipolar filaments or spindles that lack plasmid attachment will be destroyed through the dynamic instability of ParM. This is in contrast to earlier ideas in which all sister plasmids would be segregated through one bundle of filaments, containing double the number of unipolar filaments as the copy number of the plasmid in the cell<sup>19</sup>.

the high- and medium-copy-number plasmid cases, they were not observed in the low-copy-number (mini-R1) case, where partitioning via ParMRC is required for plasmid stability<sup>22</sup>. These cryo-ET data are in line with previous immuno-light microscopy data, where single pole-to-pole filaments were only observed in 40% of cells<sup>1,10</sup> and the other cells showed several localized clusters or more complex patterns.

The above data showed that ParM doublets are found in cells containing the ParMRC locus, and are probably the machinery that actively segregates plasmid DNA to opposite ends of the dividing cell, even though antiparallel arrangement of ParM filaments in cellular doublets can only be inferred from the *in vitro* studies above. It is

interesting to observe that the ratios of doublets observed per cell were the same as the ratios of the expected copy numbers of the three plasmids, although it needs to be noted that numbers remain small because of the low-throughput nature of cryo-ET (Extended Data Table 3). The ratios might indicate that each doublet carries a defined payload of DNA cargo, a fixed number of plasmids containing the *parC* locus. We propose that ParMRC spindles consisting solely of doublets elegantly circumvent the problem of synchronizing plasmid replication, filament attachment and bundle formation for all plasmids in the cell: each pair of plasmid sisters is segregated by their own spindle. The resulting asynchronous plasmid segregation is schematically summarized in Fig. 4e. Indeed, it is known that R1 plasmids are replicated randomly throughout the cell cycle<sup>23,24</sup>. In contrast, eukaryotic DNA segregation requires cohesion, kinetochore checkpoints and other dedicated machinery since all material is segregated with one coordinated and synchronized spindle.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 14 November 2014; accepted 2 March 2015.

Published online 27 April 2015.

- Møller-Jensen, J., Jensen, R. B., Löwe, J. & Gerdes, K. Prokaryotic DNA segregation by an actin-like filament. *EMBO J.* **21**, 3119–3127 (2002).
- Gerdes, K., Howard, M. & Szardenings, F. Pushing and pulling in prokaryotic DNA segregation. *Cell* **141**, 927–942 (2010).
- Gayathri, P. *et al.* A bipolar spindle of antiparallel ParM filaments drives bacterial plasmid segregation. *Science* **338**, 1334–1337 (2012).
- Orlova, A. *et al.* The structure of bacterial ParM filaments. *Nature Struct. Mol. Biol.* **14**, 921–926 (2007).
- Popp, D. *et al.* Molecular structure of the ParM polymer and the mechanism leading to its nucleotide-driven dynamic instability. *EMBO J.* **27**, 570–579 (2008).
- van den Ent, F., Møller-Jensen, J., Amos, L. A., Gerdes, K. & Löwe, J. F-actin-like filaments formed by plasmid segregation protein ParM. *EMBO J.* **21**, 6935–6943 (2002).
- Møller-Jensen, J., Ringgaard, S., Mercogliano, C. P., Gerdes, K. & Löwe, J. Structural analysis of the ParR/*parC* plasmid partition complex. *EMBO J.* **26**, 4413–4422 (2007).
- Schumacher, M. A. *et al.* Segosome structure revealed by a complex of ParR with centromere DNA. *Nature* **450**, 1268–1271 (2007).
- Garner, E. C., Campbell, C. S., Weibel, D. B. & Mullins, R. D. Reconstitution of DNA segregation driven by assembly of a prokaryotic actin homolog. *Science* **315**, 1270–1274 (2007).
- Møller-Jensen, J. *et al.* Bacterial mitosis: ParM of plasmid R1 moves plasmid DNA by an actin-like insertional polymerization mechanism. *Mol. Cell* **12**, 1477–1487 (2003).
- Izore, T., Duman, R., Kureisaite-Ciziene, D. & Löwe, J. Crenactin from *Pyrobaculum caldifontis* is closely related to actin in structure and forms steep helical filaments. *FEBS Lett.* **588**, 776–782 (2014).
- Ozyamak, E., Kollman, J., Agard, D. A. & Komeili, A. The bacterial actin MamK: *in vitro* assembly behavior and filament architecture. *J. Biol. Chem.* **288**, 4265–4277 (2013).
- Ozyamak, E., Kollman, J. M. & Komeili, A. Bacterial actins and their diversity. *Biochemistry* **52**, 6928–6939 (2013).
- Fujii, T., Iwane, A. H., Yanagida, T. & Namba, K. Direct visualization of secondary structures of F-actin by electron cryomicroscopy. *Nature* **467**, 724–728 (2010).
- von der Ecken, J. *et al.* Structure of the F-actin–tropomyosin complex. *Nature* **519**, 114–117 (2015).
- Galkin, V. E., Orlova, A., Vos, M. R., Schröder, G. F. & Egelman, E. H. Near-atomic resolution for one state of F-actin. *Structure* **23**, 173–182 (2015).
- Alushin, G. M. *et al.* High-resolution microtubule structures reveal the structural transitions in  $\alpha$ -tubulin upon GTP hydrolysis. *Cell* **157**, 1117–1129 (2014).
- Garner, E. C., Campbell, C. S. & Mullins, R. D. Dynamic instability in a DNA-segregating prokaryotic actin homolog. *Science* **306**, 1021–1025 (2004).
- Salje, J., Zuber, B. & Löwe, J. Electron cryomicroscopy of *E. coli* reveals filament bundles involved in plasmid DNA segregation. *Science* **323**, 509–512 (2009).
- Popp, D., Narita, A., Iwasa, M., Maeda, Y. & Robinson, R. C. Molecular mechanism of bundle formation by the bacterial actin ParM. *Biochem. Biophys. Res. Commun.* **391**, 1598–1603 (2010).
- Dam, M. & Gerdes, K. Partitioning of plasmid R1. Ten direct repeats flanking the *parA* promoter constitute a centromere-like partition site *parC*, that expresses incompatibility. *J. Mol. Biol.* **236**, 1289–1298 (1994).
- Breuner, A., Jensen, R. B., Dam, M., Pedersen, S. & Gerdes, K. The centromere-like *parC* locus of plasmid R1. *Mol. Microbiol.* **20**, 581–592 (1996).
- Gustafsson, P. & Nordström, K. Control of plasmid R1 replication: kinetics of replication in shifts between different copy number levels. *J. Bacteriol.* **141**, 106–110 (1980).

24. Nordström, K. Plasmid R1-replication and its control. *Plasmid* **55**, 1–26 (2006).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank F. van den Ent, K. Gerdes and P. Gayathri for help with sample preparation, and C. Johnson, C. Savva and F. de Haas for help with data collection. This work was supported by the Medical Research Council (U105184326) and the Wellcome Trust (095514/Z/11/Z). T.A.M.B. is the recipient of Federation of European Biochemical Societies (FEBS) and European Molecular Biology Organization (EMBO) (ALTF 3-2013) long-term fellowships. G.N.M. was funded by Medical Research Council grant MC-UP-A025-1012.

**Author Contributions** T.A.M.B. and J.L. designed experiments; T.A.M.B. performed experiments; T.A.M.B., G.N.M., C.S. and J.L. analysed data; T.A.M.B. and J.L. wrote the paper.

**Author Information** Cryo-EM and cryo-ET data have been deposited in the Electron Microscopy Data Bank under accession codes EMD-2848, EMD-2849 and EMD-2850. Atomic coordinates of the ParM+AMPPNP filament structure and the ParM antiparallel doublet model have been deposited in the Protein Data Bank under accession codes 5AEY and 5AI7. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.L. ([jyl@mrc-lmb.cam.ac.uk](mailto:jyl@mrc-lmb.cam.ac.uk)).

## METHODS

No statistical methods were used to predetermine sample size.

**Protein expression and purification.** ParM (UniProt: PARM\_ECOLX) and ParM mutants were expressed from plasmid *pSC1* and its derivatives<sup>25</sup> in *E. coli* BL21-AI cells and purified as described previously<sup>25</sup>. Buffer MR was used in all experiments: 50 mM Tris-HCl, 100 mM KCl, and 1 mM MgCl<sub>2</sub>, pH 7.0.

Wild-type ParM and ParM(S19R, G21R). ParM was purified by ammonium sulphate precipitation (at a final concentration of 10% (sat.) ammonium sulphate) of the lysate, followed by addition of ATP to the re-suspended pellet. ParM filaments were pelleted by centrifugation at 100,000g, and the resulting pellet containing pure protein was re-suspended in buffer and further purified by size exclusion chromatography on a Sephacryl S-200 column (GE Healthcare).

ParM(K258D, R262D). The protein was purified using a 5 ml HiTrap Q HP column (GE Healthcare), and eluted at increasing KCl concentrations. Fractions containing ParM were pooled and further purified by size exclusion on a Sephacryl S-200 column (GE Healthcare) into buffer MR. Concentrated aliquots of pure protein were frozen and stored at  $-80^{\circ}\text{C}$  until further investigation.

**Sample preparation for microscopy.** ParM+AMPPNP and ParM+ATP. ParM protein (10  $\mu\text{M}$ ) was incubated with 2 mM nucleotide in a total volume of 100  $\mu\text{l}$  for 5 min at room temperature (22  $^{\circ}\text{C}$ ) before cryo-EM sample preparation. The same procedure was used for polymerization of the ParM S19R, G21R mutant.

ParM+ATP+vanadate and ParM+ADP+vanadate. ParM protein (10  $\mu\text{M}$ ) was incubated with 2 mM ATP or ADP and 4 mM sodium orthovanadate in a total volume of 100  $\mu\text{l}$  for 2 h or 5 min at room temperature before cryo-EM sample preparation.

ParM+ADP. 400  $\mu\text{M}$  ParM was incubated with 10 mM ADP in a total volume of 25  $\mu\text{l}$  for 5 min at room temperature.

ParM(K258D, R262D)+AMPPNP. Protein (60  $\mu\text{M}$ ) was incubated with 2 mM AMPPNP in a total volume of 100  $\mu\text{l}$  for 5 min at room temperature.

ParM *in vitro* doublets. ParM protein (20  $\mu\text{M}$ ) was incubated with 2 mM AMPPNP in the presence of 2% (w/v) PEG 6000 in a 100  $\mu\text{l}$  for 5 min at room temperature.

ParM(D170A)-overexpressing cells. ParM was expressed to high levels for cryo-ET using the plasmid pRBJ212 (ParM(D170A), *ptac* promoter)<sup>1</sup> transformed into *E. coli* B/R266 cells. Cells were grown in M9 medium at 30  $^{\circ}\text{C}$  and induced with 1 mM IPTG at an attenuation  $D_{600\text{ nm}} \approx 0.5$ . Samples were prepared 4 h after induction. Bacterial cells with different copy-number plasmids containing the ParMRC locus. *E. coli* strain B/R266 (ref. 26) was transformed with high- (pDD19), medium- (pKG321) or low-copy (pKG491) plasmids and grown in M9 medium supplemented with 100  $\mu\text{g ml}^{-1}$  ampicillin at 30  $^{\circ}\text{C}$  (ref. 21). Cells were grown to  $D_{600\text{ nm}} \approx 0.4\text{--}0.6$  (grown to logarithmic growth phase) before sample preparation for cryo-EM.

**Cryo-EM grid preparation.** Samples for cryo-EM were prepared by pipetting 2.5  $\mu\text{l}$  of the sample onto a freshly glow-discharged Quantifoil Cu/Rh 200 mesh grids (R2/2 for purified protein, and R3.5/1 for cellular tomography) and plunge frozen into liquid ethane in a Vitrobot Mark IV (FEI). Only for cryo-ET, 11  $\mu\text{l}$  of sample were pre-mixed with 1  $\mu\text{l}$  of protein-A conjugated with 10 nm colloidal gold (Cell Microscopy Center, Utrecht University, The Netherlands). Plunged grids were transferred to liquid nitrogen and stored.

**Electron microscopy data collection.** Two-dimensional cryo-EM data were collected using either an FEI Krios microscope operated at 300 kV or a FEI Spirit microscope operated at 120 kV. High-throughput data were collected on the FEI Krios using EPU software at an unbinned calibrated pixel size of 1.30  $\text{\AA}$  or 1.07  $\text{\AA}$  on a Falcon II direct electron detector. A combined total dose of 25–32 electrons per square  $\text{\AA}$  was applied with each exposure that lasted 1 s. Images were collected at 1–6  $\mu\text{m}$  underfocus. Tilt series data were collected on an FEI Krios equipped with a Quantum energy filter (Gatan) using SerialEM software<sup>27</sup>, on a K2 direct electron detector operating in counting mode. Tilt series data were typically collected from  $\pm 60^{\circ}$  with  $1^{\circ}$  tilt increment at 4–12  $\mu\text{m}$  underfocus with a combined dose of about 120 electrons per square  $\text{\AA}$  applied over the entire series.

**Image processing and data analysis.** Real-space helical reconstruction. An averaged power spectrum for each cryo-EM image was calculated using CTFFIND<sup>28</sup>, and images showing clear Thon rings were retained. ParM filaments were extracted from the selected images using SPRING and EMAN2<sup>29,30</sup>. The helical symmetry of each sample was accurately determined by comparing the power spectrum of the aligned segments with power spectra of re-projections of the calculated reconstructions. Experimentally determined helical parameters (Extended Data Table 1) were used for refinement using the program segmentrefine3D in SPRING. The final volumes were compensated for the B-factor and filtered to the obtained resolutions (Extended Data Table 1). Resolution of the structure was estimated using gold-standard Fourier shell correlation measurements in SPRING and additionally using ResMap<sup>31</sup>. Visualization of densities was performed in UCSF Chimera<sup>32</sup>.

Atomic model building. The atomic structure from PDB 4A62 (ref. 3) was fitted into the cryo-EM density of ParM+AMPPNP using MolRep<sup>33</sup>. Maximum-likelihood refinement of the atomic structure against the cryo-EM density was performed in REFMAC5 (ref. 34) using standard protein stereochemistry and additional external restraints based on PDB 4A62, generated in ProSMART<sup>35</sup>. Model building was performed in COOT and MAIN<sup>36,37</sup>.

Rigid body fitting. ParM was divided into two sub-domains, based on the previous ParM+ADP X-ray structure (PDB 1MWM). Each sub-domain was declared as a rigid body and these were fitted into the ParM+ADP filament structure using REFMAC5.

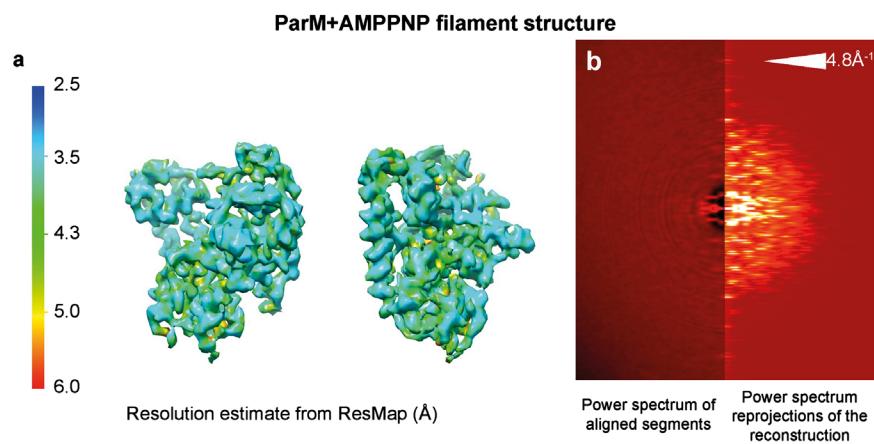
Polarity assignment of ParM filaments in doublets. First, images of ParM doublets were carefully selected on the basis of image quality (as assessed by a visual inspection of power spectra), and by a visual assessment of the distance between the two filaments in the doublet. The assumption made from the appearance of the class averages was that images in which the distance between the centres of two ParM filaments in the doublet was maximum would show ParM filaments entirely in the same  $x$ - $y$  plane (the  $z$ -axis being the path of the electron beam in the microscope). The two ParM filaments in all the doublets in these selected images were picked manually using EMAN2 (ref. 30). The manual pick was used to extract short segments along each filament in the doublet. The extracted segments were aligned to re-projections of the high-resolution ParM+AMPPNP filament model using SPRING<sup>29</sup>. In five out of the six doublets analysed, the assigned directionality of ParM filaments was predominantly anti-parallel and in one case the assignment was mostly parallel.

Derivation of the doublet model. The ParM doublet is not a true helical specimen, thus conventional helical reconstruction could not be performed. This difficulty of characterizing higher-order filament structures of ParM filaments has been previously reported<sup>20</sup>. The average distance between the centres of the two ParM filaments in the doublet was found to be 65.1  $\text{\AA}$  by analysis of one-dimensional line-profiles of all obtained class averages. Two copies of the high-resolution cryo-EM structure of the ParM+AMPPNP filaments were accordingly placed with their centres 65.1  $\text{\AA}$  apart in space in an anti-parallel orientation. The placement was repeated for all possible combinations of the azimuthal angles of both filaments. Re-projections of all these resulting volumes were aligned with all obtained class averages. As expected intuitively from an inspection of the class averages, models in which the thickest part of one ParM filament overlapped with the thinnest part of the other filament in the doublet had higher cross-correlation scores. We placed two copies of the atomic structure of the ParM+AMPPNP in the volume with the highest score. Since this was not a standard cryo-EM reconstruction, meaning resulting atomic accuracy would be somewhat lower, we only used the  $\text{C}\alpha$  atoms for determining distances shown in Extended Data Table 2.

Tomographic reconstructions. Tilt series data were aligned using IMOD<sup>38</sup> and three-dimensional reconstructions were conducted using the SIRT algorithm implemented in Tomo3D<sup>39</sup>. Visualization of data was done using IMOD and UCSF Chimera<sup>32</sup>.

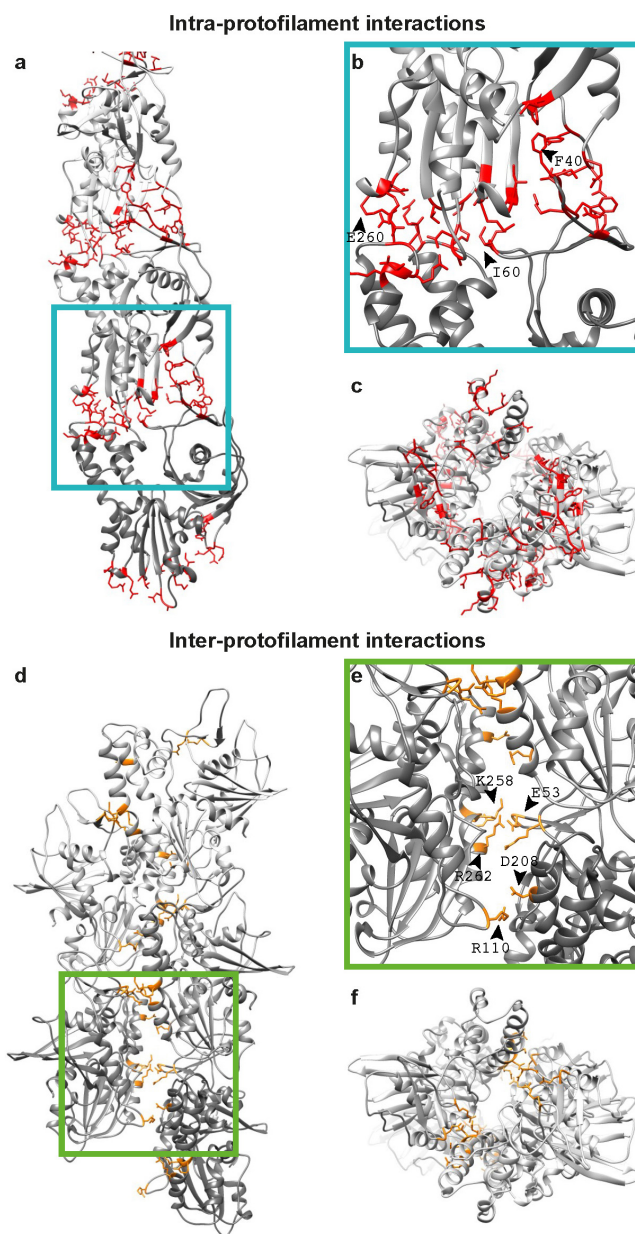
25. Salje, J. & Löwe, J. Bacterial actin: architecture of the ParMRC plasmid DNA partitioning complex. *EMBO J.* **27**, 2230–2238 (2008).
26. Cooper, S. & Helmstetter, C. E. Chromosome replication and the division cycle of *Escherichia coli* B/r. *J. Mol. Biol.* **31**, 519–540 (1968).
27. Mastrorade, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).
28. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347 (2003).
29. Desfosses, A., Ciuffa, R., Gutsche, I. & Sachse, C. SPRING – an image processing package for single-particle based helical reconstruction from electron cryomicrographs. *J. Struct. Biol.* **185**, 15–26 (2014).
30. Tang, G. *et al.* EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).
31. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nature Methods* **11**, 63–65 (2014).
32. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
33. Vagin, A. & Teplov, A. Molecular replacement with MOLREP. *Acta Crystallogr. D* **66**, 22–25 (2010).
34. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67**, 355–367 (2011).
35. Nicholls, R. A., Fischer, M., McNicholas, S. & Murshudov, G. N. Conformation-independent structural comparison of macromolecules with ProSMART. *Acta Crystallogr. D* **70**, 2487–2499 (2014).
36. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
37. Turk, D. MAIN software for density averaging, model building, structure refinement and validation. *Acta Crystallogr. D* **69**, 1342–1357 (2013).
38. Kremer, J. R., Mastrorade, D. N. & McIntosh, J. R. Computer visualization of three-dimensional image data using IMOD. *J. Struct. Biol.* **116**, 71–76 (1996).
39. Agulleiro, J. I. & Fernandez, J. J. Fast tomographic reconstruction on multicore computers. *Bioinformatics* **27**, 582–583 (2011).





**Extended Data Figure 1 | Resolution estimate of the ParM+AMPPNP reconstruction.** **a**, Resolution of the ParM+AMPPNP reconstruction was estimated using ResMap and this estimate was plotted back onto the cryo-EM density. Blue indicates high resolution; red indicates lower resolution. **b**, The

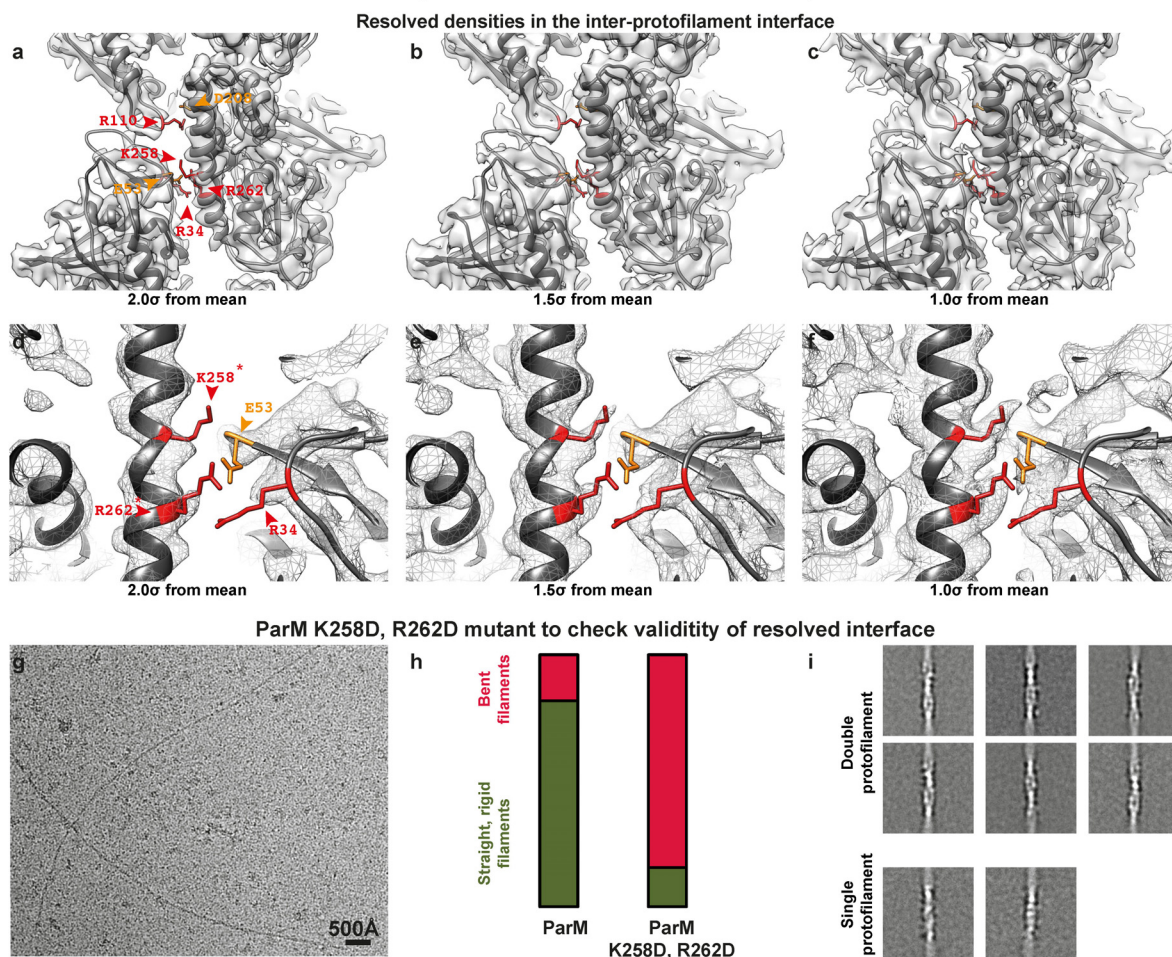
power spectrum of the aligned segments (left) compared with the power spectrum of the re-projection of the cryo-EM reconstruction (right). A reflection is observed in both cases at  $4.8 \text{ Å}^{-1}$ , indicating that the resolution extends beyond this shell. See Fig. 2e for Fourier shell correlation curves.



**Extended Data Figure 2 | Intra- and inter-protofilament interactions in ParM filaments.** **a**, Atomic model of one protofilament (strand) of ParM is shown with the residues at the protein–protein interface highlighted in red. See Extended Data Table 2 for a detailed list. **b**, A magnified view of the interface. Three residues at the interface have been labelled. **c**, The complete ParM

filament (that is, both protofilaments/strands) shown end-on. **d**, Atomic model of the ParM filament with the inter-protofilament residues at the protein–protein interface highlighted in orange. **e**, A magnified view of **d**. Salt bridging residues are labelled. **f**, An orthogonal view of **d**. See Extended Data Table 2 for a detailed list of interacting residues.

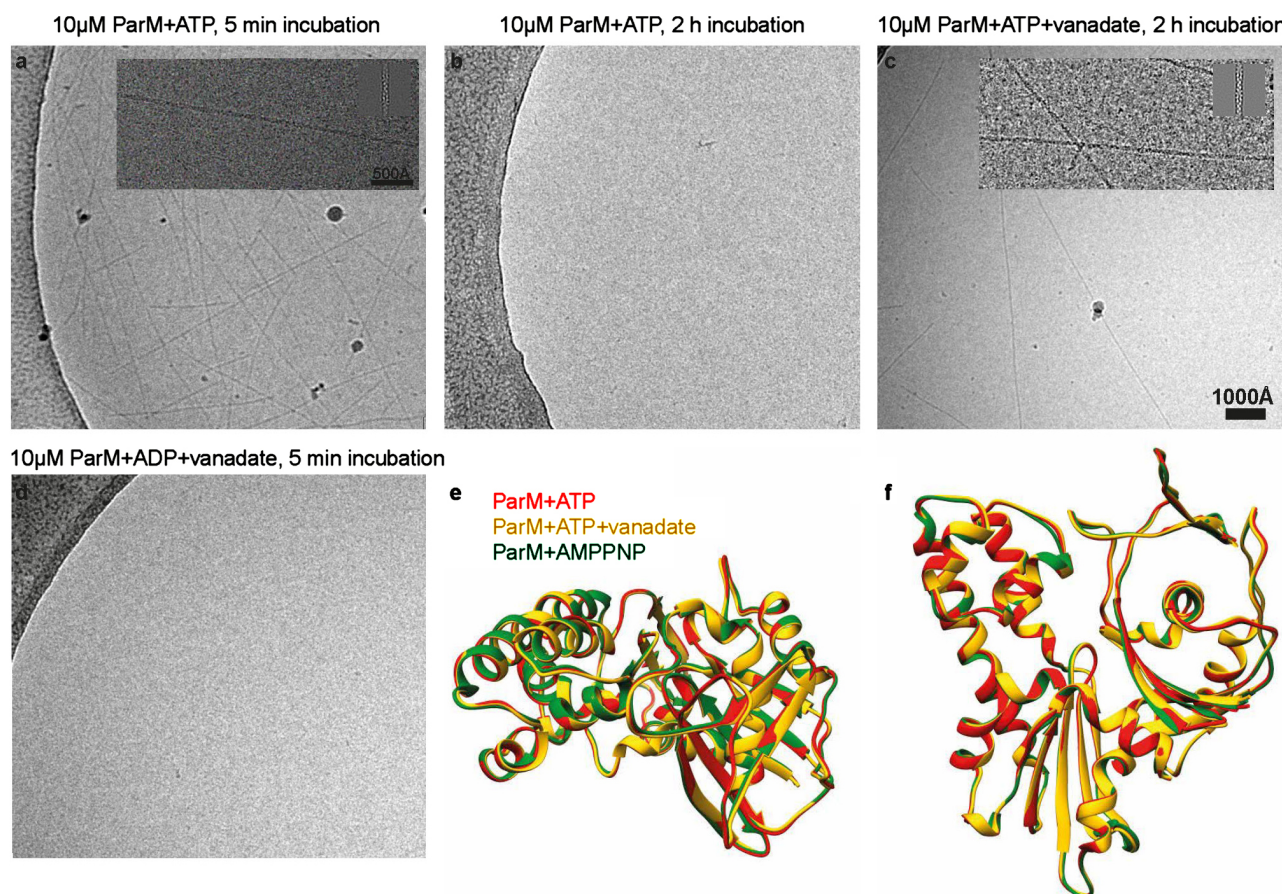
## ParM inter-protofilament interface is small but important



**Extended Data Figure 3 | The ParM inter-protofilament interface is small but important.** **a**, Cryo-EM density for the ParM+AMPPNP filament is shown at an isosurface contour level of  $2.0\sigma$  from the mean value. Overlaid on the density, refined atomic coordinates from REFMAC are additionally displayed as grey ribbons. Residues forming salt bridges at the inter-protofilament interface are highlighted. **b**, The same figure as **a**, except the cryo-EM density shown at an isosurface contour level of  $1.5\sigma$  from the mean; **c**,  $1.0\sigma$  from the mean. **d–f**, A magnified view of the primary salt-bridged interface consisting of charged residues that form the ParM inter-protofilament interface. The cryo-EM density is shown as a mesh at three different contour levels to demonstrate resolved side-chain densities. Positively charged residues are highlighted in red; negatively charged residues are highlighted in orange. **g**, Two residues (K258 and R262) that were the best resolved (marked with an asterisk in **d**), were mutated to aspartic acid to test the importance of this inter-protofilament interface. A cryo-EM image of this mutant protein assembled with AMPPNP is shown. A much higher concentration of the protein was

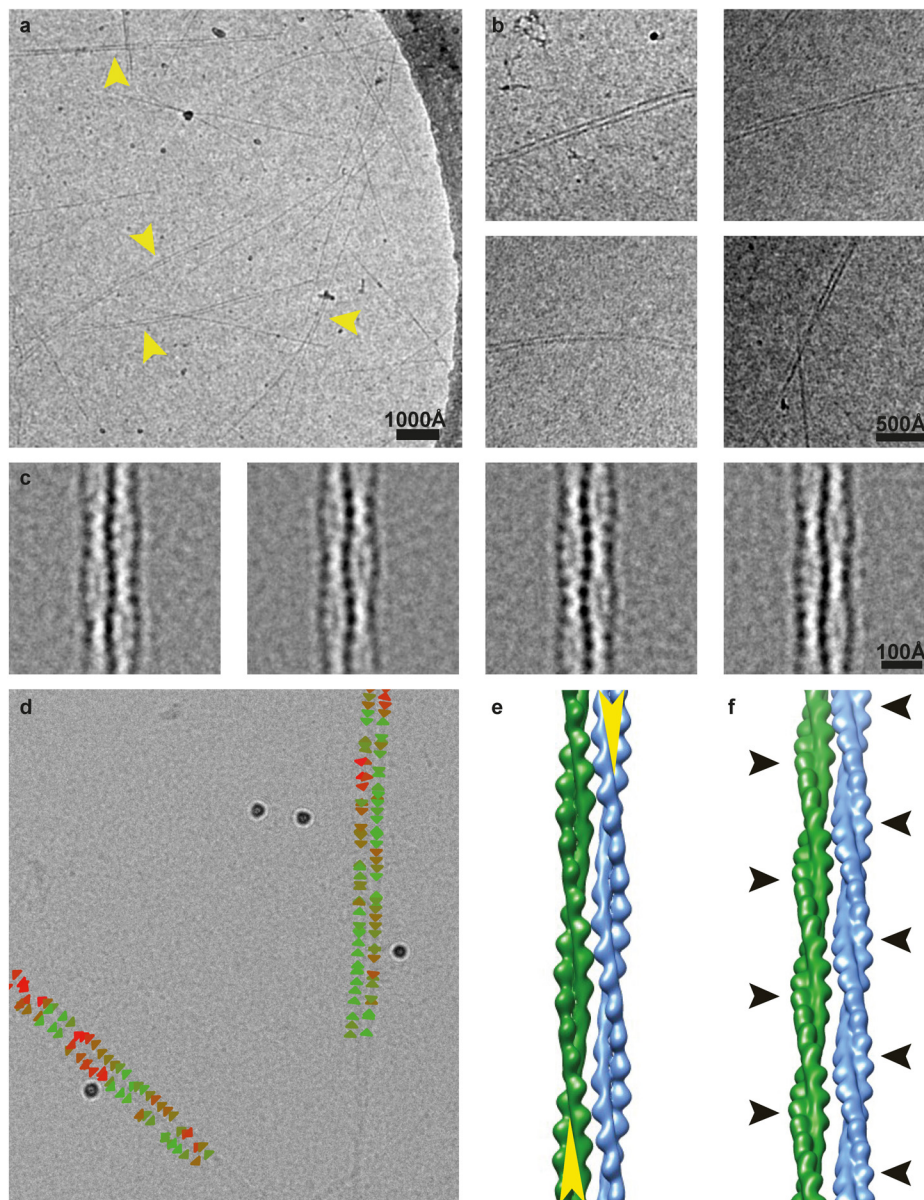
required to obtain filaments on cryo-EM grids (Methods). This experiment was repeated four times. **h**, Randomly selected cryo-EM images of ParM+AMPPNP and ParM(K258D, R262D)+AMPPNP were used to count occurrences of straight and bent filaments by visual inspection. The results of this quantification are shown as a percentage bar diagram. For the ParM protein, 82% of all filaments were classified as straight, while 18% were bent ( $n = 345$ ). Using exactly the same classification criteria, only 15% of the filaments were found to be straight and 85% of the filaments were bent ( $n = 45$ ) for the ParM(K258D, R262D) mutant protein. **i**, Reference-free class averages show that most of the ParM(K258D, R262D) filaments are made up of double protofilaments like wild-type ParM. Some class averages show evidence of bending. A few class averages show that single protofilaments were present in the sample (lower panels). However, the double mutation destabilizes the entire ParM filament, making filament formation an unfavourable reaction, illustrating that even though the inter-protofilament interface is small, it is critical for ParM filament formation.





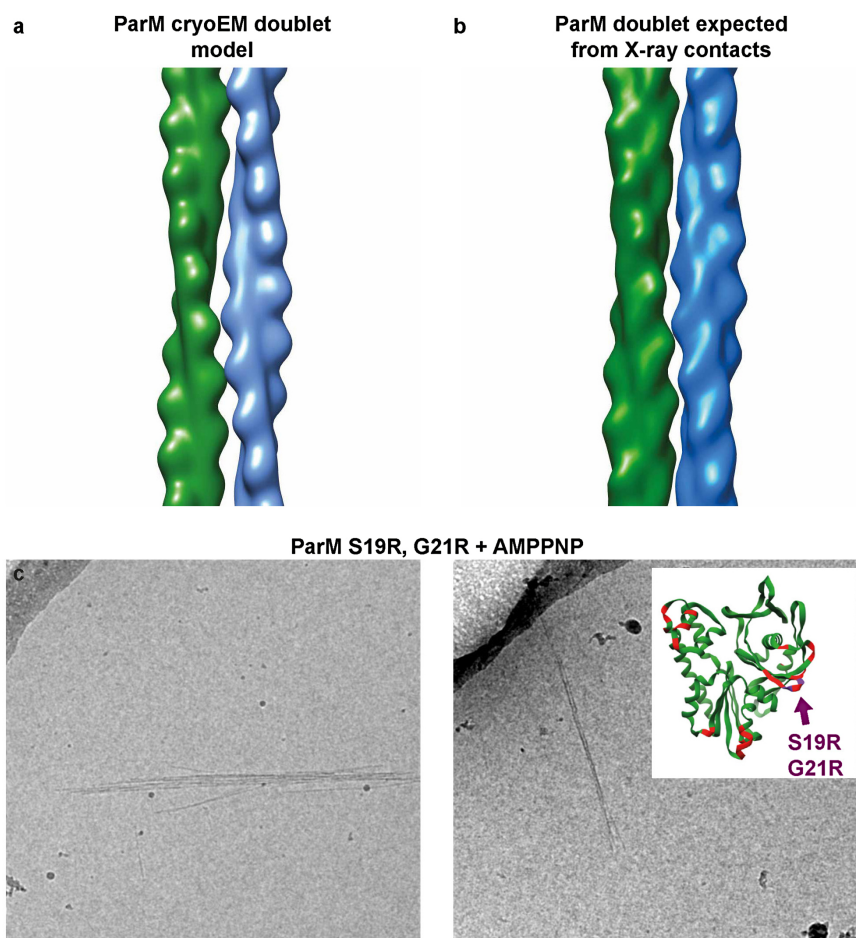
**Extended Data Figure 4 | ParM adopts a compact conformation until ATP is hydrolysed to ADP or until phosphate is released.** **a**, ParM protein (10  $\mu$ M) was incubated with ATP (2 mM) and cryo-EM samples were prepared after 5 min. Many filaments were observed on the grid. This experiment was repeated ten times. **b**, After 2 h, no filaments were seen in the same reaction. Presumably, ATP had been hydrolysed and ParM had returned to monomeric form. This experiment was repeated three times. **c**, When sodium orthovanadate (4 mM) was included in the reaction, filaments could be observed, even after 2 h. This experiment was repeated three times. **d**, The same

reaction as **a**, except ATP was replaced by ADP. No filaments were observed in this reaction. This experiment was repeated four times. **e**, **f**, We performed real-space helical reconstruction of the ParM+ATP filaments (red) and ParM+ATP+vanadate filaments (yellow), and compared them with the ParM+AMPPNP filament structure (green). Comparison shows that ParM is held in a very similar conformation until hydrolysis of ATP is complete or until phosphate is released since we currently cannot distinguish these two possible effects of vanadate. See Fig. 2e for resolution estimates and Extended Data Table 1 for image-processing statistics.



**Extended Data Figure 5 | Model of the ParM doublet.** **a**, A cryo-EM image of ParM+AMPPNP + 2% PEG 6000. Instances of doublets are marked with yellow arrowheads. This experiment was repeated 15 times. **b**, More examples of ParM doublets observed in cryo-EM. **c**, Class averages of the doublets. **d**, Directionality assignment of the filaments in the doublet. Individual sub-segments and their assigned directionality are indicated by

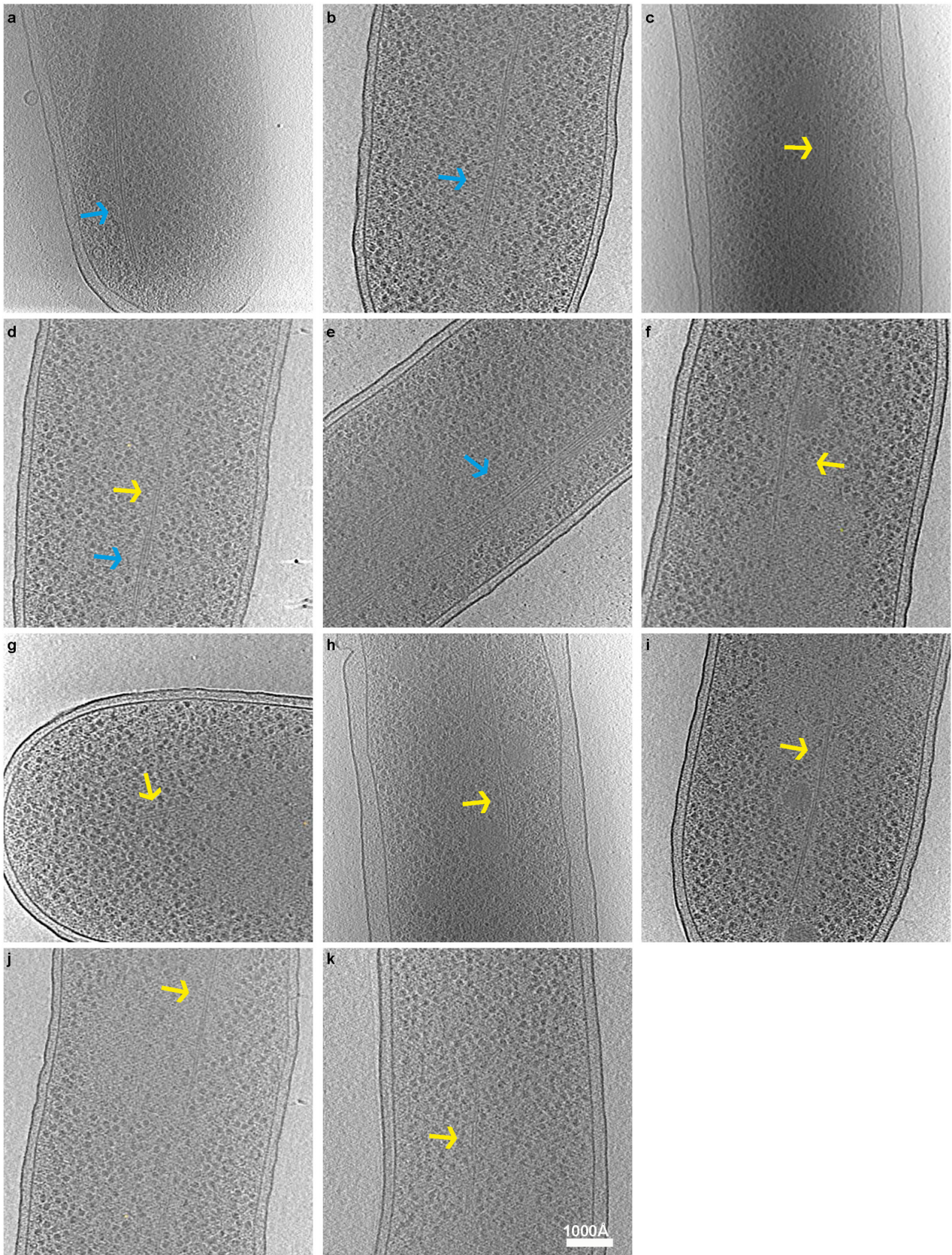
triangles, coloured on the basis of the cross-correlation score in the alignment procedure: red indicates a poor cross correlation score; green indicates a good score. **e**, A schematic model of the anti-parallel ParM doublet. Directionality is indicated with a yellow arrow. **f**, The thickest parts of ParM filaments of the doublet (as they appear in projection) are marked with black arrowheads.



**Extended Data Figure 6 | Validation of the doublet model.** **a**, Two ParM filaments arranged in an anti-parallel orientation, as obtained from the ParM cryo-EM doublet model. **b**, Two ParM filaments arranged in an anti-parallel orientation, obtained from crystal packing of a monomeric ParM X-ray structure (PDB 4A62)<sup>3</sup>. **c**, Two residues at the interface of the doublet (see

Extended Data Table 2), S19 and G21, were mutated to arginine to improve affinity of ParM filaments to each other. Cryo-EM images of the mutant protein with AMPPNP show spontaneous doublet formation and filament bundling without crowding agent, validating the doublet model. This experiment was repeated six times.





**Extended Data Figure 7 | ParM bundles and doublets observed *in vivo*.** **a–k,** *E. coli* B/R266 cells were transformed with a high-copy (pDD19) or medium-copy (pKG321) plasmid containing the ParMRC locus. Transformed cells were grown to log phase and then prepared for cryo-EM. This figure shows a gallery of ParM bundles (blue arrows) and doublets

(yellow arrows) observed in these cells. **a, c, e, h,** Cells transformed with the high-copy-number plasmid; **b, d, f, g, i, j, k,** cells transformed with the medium-copy-number plasmid. Each experiment with different copy-number plasmids was performed only once owing to the low-throughput nature of cryo-ET.

**Extended Data Table 1 | Image processing statistics for cryo-EM reconstructions of ParM filaments**

|  | ParM+AMPPNP | ParM+ATP | ParM+ATP+<br>vanadate | ParM+ADP |
|--|-------------|----------|-----------------------|----------|
| Resolution, FSC at 0.143 (Å)                 | 4.3         | 7.5      | 6.4                   | 11.0     |
| Filament pitch (Å)                           | 54.0        | 54.0     | 53.8                  | 51.0     |
| Subunits per turn                            | 2.18        | 2.18     | 2.18                  | 2.18     |
| Input segment step size (Å)                  | 268         | 161      | 161                   | 153      |
| Segment size for alignment (Å <sup>2</sup> ) | 364x364     | 400x400  | 400x400               | 700x700  |
| Asymmetric units included                    | 561,231     | 13,825   | 122,864               | 53,648   |

Extended Data Table 2 | Interfaces forming residues of ParM

| Interface                          | Residue Numbers | Sequence   |
|------------------------------------|-----------------|--|
| Inter-protofilament interface      | 1               | MLVFIDDGSTNIKLQWQESDGTIKQHISPNSFKREWAVSFGDKKVFNYTLNGEQYSFDPI |
|                                    | 61              | SPDAVVTNIAWQYSDVNVVAVHHALLTSGLPVSEVDIVCTLPLTEYDRNNQPNTENIE   |
|                                    | 121             | RKKANFRKKITLNGGDTFTIKDVKVPESIPAGYEVQLQELDELSLLIIDLGGTTLDISQ  |
|                                    | 181             | VMGKLSGISKIYGDSSLGVSLVTSAVKDALSARTKGSSYLADDIIHRKDNNYLKQRIN   |
|                                    | 241             | DENKISIVTEAMNEALRKLEQRLVNTLNEFSGYTHVMVIGGGAELICDAVKKHTQIRDER |
|                                    | 301             | FFKTNNSQYDLVNGMYLIGN   |
| Intra-protofilament interface      | 1               | MLVFIDDGSTNIKLQWQESDGTIKQHISPNSFKREWAVSFGDKKVFNYTLNGEQYSFDPI |
|                                    | 61              | SPDAVVTNIAWQYSDVNVVAVHHALLTSGLPVSEVDIVCTLPLTEYDRNNQPNTENIE   |
|                                    | 121             | RKKANFRKKITLNGGDTFTIKDVKVPESIPAGYEVQLQELDELSLLIIDLGGTTLDISQ  |
|                                    | 181             | VMGKLSGISK IYGDSSLGVSLVTSAVKDALSARTKGSSYLADDIIHRKDNNYLKQRIN  |
|                                    | 241             | DENKISIVTEAMNEALRKLEQRLVNTLNEFSGYTHVMVIGGGAELICDAVKKHTQIRDER |
|                                    | 301             | FFKTNNSQYDLVNGMYLIGN   |
| Inter-filament interface (doublet) | 1               | MLVFIDDGSTNIKLQWQESDGTIKQHISPNSFKREWAVSFGDKKVFNYTLNGEQYSFDPI |
|                                    | 61              | SPDAVVTNIAWQYSDVNVVAVHHALLTSGLPVSEVDIVCTLPLTEYDRNNQPNTENIE   |
|                                    | 121             | RKKANFRKKITLNGGDTFTIKDVKVPESIPAGYEVQLQELDELSLLIIDLGGTTLDISQ  |
|                                    | 181             | VMGKLSGISKIYGDSSLGVSLVTSAVKDALSARTKGSSYLADDIIHRKDNNYLKQRIN   |
|                                    | 241             | DENKISIVTEAMNEALRKLEQRLVNTLNEFSGYTHVMVIGGGAELICDAVKKHTQIRDER |
|                                    | 301             | FFKTNNSQYDLVNGMYLIGN   |

Residues of ParM that are part of interfaces have been highlighted. Inter- and intra-protofilament interface-forming residues have been highlighted in red and green respectively. These residues have been assigned using a 4 Å distance cut-off based on the ParM+AMPPNP structure. Residues forming the inter-filament interface in the ParM doublet have been highlighted in blue. This assignment was based on a 7 Å distance cut-off for C $\alpha$  atoms in the derived model of the ParM doublet because of the lower accuracy of the model.



**Extended Data Table 3 | Instances of single, double and bundled ParM filaments seen in bacterial cells with different copy-number plasmids**

| Plasmid type | Single filaments | Double filaments | Bundles | Total number of cells imaged | Doublets per cell |
|--------------|------------------|------------------|---------|------------------------------|-------------------|
| High copy    | 5                | 35               | 8       | 6                            | 5.83              |
| Medium copy  | 11               | 36               | 2       | 23                           | 1.56              |
| Low copy     | 4                | 4                | 0       | 14                           | 0.28              |

The ParMRC locus was inserted into high-, medium- and low-copy-number plasmids (plasmids pDD19, pKG321 and pKG491 respectively). These plasmids were in turn inserted into *E. coli* cells and imaged using cryo-ET. The ratio of observed doublets per cell (that is, the number of doublets observed divided by the number of cells imaged) was 5.8:1.6:0.3 (~19:5:1). These ratios are roughly the same as the expected copy-number ratios of the different copy-number plasmids.

# Structures of human phosphofructokinase-1 and atomic basis of cancer-associated mutations

Bradley A. Webb<sup>1\*</sup>, Farhad Forouhar<sup>2\*</sup>, Fu-En Szu<sup>2</sup>, Jayaraman Seetharaman<sup>2</sup>, Liang Tong<sup>2</sup> & Diane L. Barber<sup>1</sup>

**Phosphofructokinase-1 (PFK1), the ‘gatekeeper’ of glycolysis, catalyses the committed step of the glycolytic pathway by converting fructose-6-phosphate to fructose-1,6-bisphosphate. Allosteric activation and inhibition of PFK1 by over ten metabolites and in response to hormonal signalling fine-tune glycolytic flux to meet energy requirements<sup>1</sup>. Mutations inhibiting PFK1 activity cause glycogen storage disease type VII, also known as Tarui disease<sup>2</sup>, and mice deficient in muscle PFK1 have decreased fat stores<sup>3</sup>. Additionally, PFK1 is proposed to have important roles in metabolic reprogramming in cancer<sup>4,5</sup>. Despite its critical role in glucose flux, the biologically relevant crystal structure of the mammalian PFK1 tetramer has not been determined. Here we report the first structures of the mammalian PFK1 tetramer, for the human platelet isoform (PFKP), in complex with ATP-Mg<sup>2+</sup> and ADP at 3.1 and 3.4 Å, respectively. The structures reveal substantial conformational changes in the enzyme upon nucleotide hydrolysis as well as a unique tetramer interface. Mutations of residues in this interface can affect tetramer formation, enzyme catalysis and regulation, indicating the functional importance of the tetramer. With altered glycolytic flux being a hallmark of cancers<sup>6</sup>, these new structures allow a molecular understanding of the functional consequences of somatic PFK1 mutations identified in human cancers. We characterize three of these mutations and show they have distinct effects on allosteric regulation of PFKP activity and lactate production. The PFKP structural blueprint for somatic mutations as well as the catalytic site can guide therapeutic targeting of PFK1 activity to control dysregulated glycolysis in disease.**

Previous attempts to obtain the structure of mammalian tetrameric PFK1 used native protein or recombinant protein generated in yeast or bacteria. A limitation of using native PFK1 is that most mammalian tissues express all three isoforms: muscle (PFKM), liver (PFKL) and platelet (PFKP)<sup>7</sup>. Although there are structures of PFK from prokaryotes<sup>8–11</sup> and eukaryotes<sup>12–14</sup>, including dimeric rabbit PFKM expressed in *Escherichia coli*<sup>12</sup>, these structures provide limited information on the catalytic interface or the conformational changes with regulation of the tetrameric mammalian enzyme. To overcome current limitations with structural studies of human PFK1, we produced recombinant PFKP by using a baculovirus expression system. The recombinant enzyme, purified to homogeneity (Extended Data Fig. 1a), is tetrameric as shown by transmission electron microscopy (TEM; Fig. 1a). The activity and regulation of recombinant PFKP, including high cooperativity for fructose-6-phosphate (F6P), a high affinity for ATP-Mg<sup>2+</sup>, and high sensitivity to ATP inhibition (Extended Data Fig. 1b, c), was similar to previously reported mouse PFKP expressed in yeast<sup>15</sup>.

We determined the crystal structure of the PFKP tetramer in complex with ATP-Mg<sup>2+</sup> at 3.1 Å resolution (Fig. 1b–d and Extended Data Fig. 2). The atomic model has good agreement with the crystallographic data and the expected geometric parameters (Extended Data Table 1). The asymmetric unit contained two tetramers, and the eight protomers have essentially the same conformation (with r.m.s.d.

of ~0.3 Å between any pair of them, Extended Data Fig. 3). The overall organizations of the two tetramers are slightly different, reflected in part by changes in the relative orientations of the two dimers (Extended Data Fig. 3).

Each PFKP tetramer measures 13.8 nm by 10.3 nm, similar in size and shape to what we calculated from TEM images (Fig. 1a, b). The tetramer is composed of a dimer of dimers, and the interface between the two dimers is relatively small, with a buried surface area of 700 Å<sup>2</sup> for each subunit (arrow labelled ‘t’ in Fig. 1b, c). The two subunits of the dimer are arranged in an antiparallel orientation, confirming previous predictions<sup>16</sup>, with a buried surface area of 1,800 Å<sup>2</sup> for each subunit. The active site is located at the interface between the two subunits (arrow labelled ‘c’ in Fig. 1b, e).

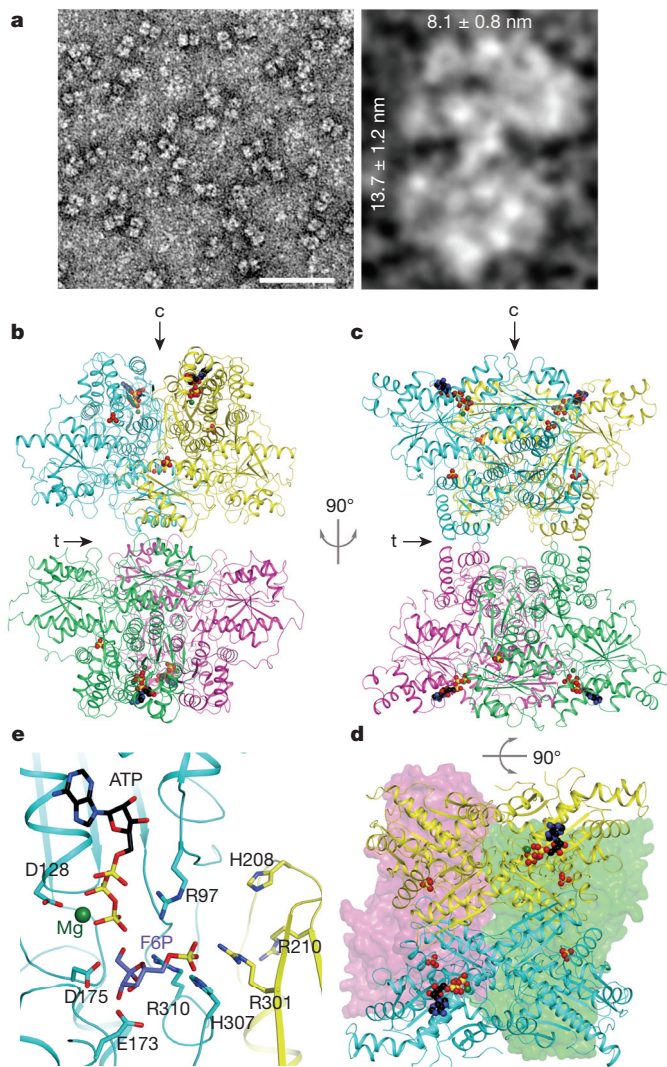
The structure of PFKP is likely to represent the active conformation of the enzyme. The crystal was prepared at pH 7, near physiological pH, and residues in the active site that are important for substrate binding and/or catalysis have similar conformations in PFKP as in other PFK structures (Fig. 1e). The F6P substrate, as observed in the *Saccharomyces cerevisiae* PFK (ScPFK) structure<sup>12</sup>, can be readily accommodated in the PFKP active site for catalysis. The invariant substrate binding residues His208 and Arg210 from the second protomer of the dimer are located ~6 Å away from F6P, suggesting that a closure of this region of the active site may occur upon F6P binding and catalysis. PFKP contains only one ATP in each subunit bound to the active site, despite the presence of 10 mM ATP during crystallization and even though the allosteric adenine nucleotide-binding sites are functional (Extended Data Fig. 1b, c). We also observed the binding of two phosphate groups in each protomer at positions corresponding to the prokaryotic PFK effector sites (Extended Data Fig. 2c, d). The enzyme activity, regulation and stability of PFK1 are controlled by binding phosphate or sulphate ions<sup>17,18</sup>. PFKP displayed a loss of ATP inhibition in the presence of 10 mM sodium sulphate (Extended Data Fig. 2e), suggesting that phosphate-binding and inhibitory site ATP-binding are mutually exclusive in the tetrameric structure.

We also determined the crystal structure of PFKP in complex with ADP at 3.4 Å resolution, at pH 7.5. The relatively low resolution of this structure precludes a detailed structural comparison with that of the ATP-Mg<sup>2+</sup> complex. However, it is clear there is a dramatic change in the relative positions of the two domains in each protomer (Fig. 2a), and especially the overall structures of the dimer and tetramer (Fig. 2b). A rotation of ~12° is observed between the subdomains of the ADP complex protomer relatively to the ATP complex, leading to an 8 Å shift in the substrate binding domain relative to the nucleotide binding domain. An effect of this conformational change is to open the catalytic site (Fig. 2c, d), which may play a role in the release of products. The conformational changes observed between the ATP and ADP complexes of PFKP are different from those seen for the R- and T-states of bacterial PFK (Extended Data Fig. 4)<sup>8,19</sup>.

We tested the importance of hydrophobic and electrostatic interactions at the tetramer interface for enzyme activity (Fig. 3a). Most

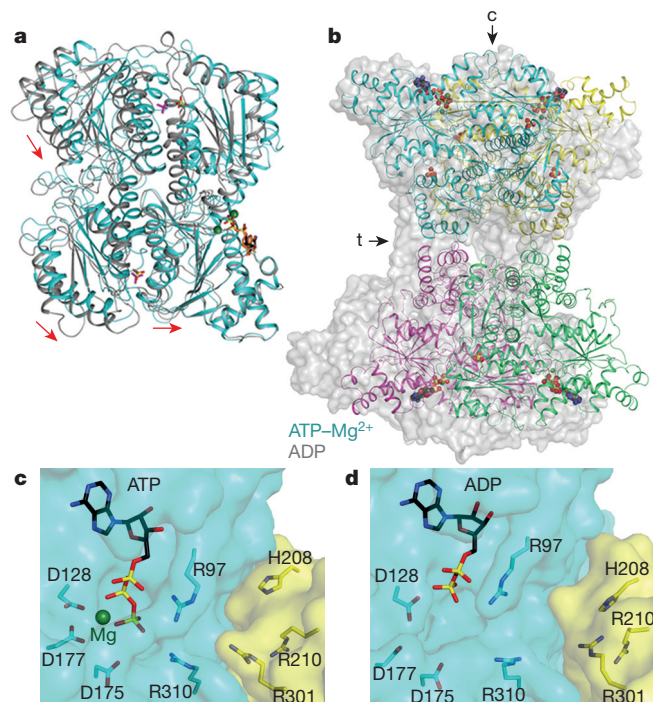
<sup>1</sup>Department of Cell and Tissue Biology, University of California, San Francisco, California 94143, USA. <sup>2</sup>Department of Biological Sciences, Northeast Structural Genomics Consortium, Columbia University, New York, New York 10027, USA.

\*These authors contributed equally to this work.



**Figure 1 | Structure of ATP-bound tetrameric PFKP is in the active conformation.** **a**, TEM images of PFKP. Left: scale bar, 50 nm. Right: indicated dimensions are the mean  $\pm$  s.d. of 37 individual particles. **b**, **c**, Ribbon diagrams of PFKP displaying the relative orientation of PFKP tetramer subunits. Each subunit is individually coloured. Arrows labelled 'c' and 't' indicate the catalytic and tetramer interfaces, respectively. View in **c** is rotated 90° from **b**. **d**, View rotated 90° from **c** displaying the catalytic sites. The front subunits are shown in ribbon representation and the rear subunits depicted as surface models. ATP, black;  $\text{Mg}^{2+}$ , dark green; phosphate, yellow. **e**, The binding mode of ATP- $\text{Mg}^{2+}$  at the active site of PFKP. The binding mode of the F6P substrate in ScPFK<sup>12</sup> is also shown.

residues at the interface are hydrophobic. Tyr645 and Phe649 from the two subunits form a  $\pi$ -stack of four aromatic side chains in the interface, with Phe649 in the middle (Fig. 3a). Despite the overall similarities in organization between the PFKP tetramer to that of ScPFK<sup>12</sup>, there are significant differences in the tetramer interface between the two enzymes (Extended Data Fig. 5). Phe649, which is evolutionarily conserved in metazoans but not in yeasts, is a leucine residue in ScPFK  $\alpha$ -subunit. However, this Leu residue has a completely different local environment in ScPFK compared with Phe649 in PFKP. We generated recombinant PFKP with Phe649 mutated to Leu (Extended Data Fig. 6a) to test whether Phe649 is required for tetramer formation. Previous studies showed that PFK1 assembles into tetramers in a concentration- and ligand-dependent manner, with allosteric activators favouring the formation of tetramers and allosteric inhibitors favouring the formation of dimers<sup>20–22</sup>. In a buffer containing ADP,



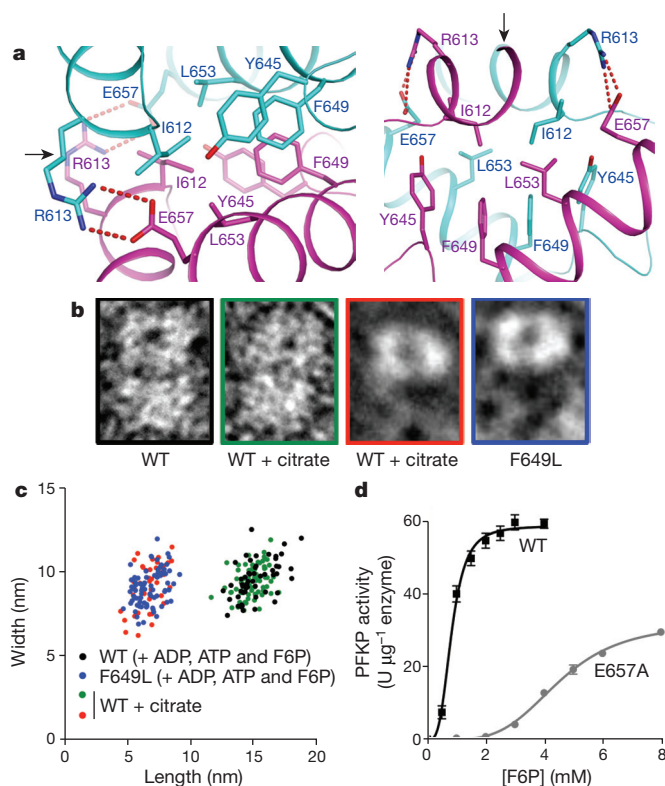
**Figure 2 | PFKP undergoes a large conformational change upon ATP hydrolysis.** **a**, Structural overlay of ATP-bound (cyan) and ADP-bound (grey) PFKP subunits. **b**, Structural overlay of ATP-bound (coloured ribbon) and ADP-bound (grey surface) PFKP tetramers. Arrows labelled 'c' and 't' represent the catalytic and tetramer interfaces, respectively. **c**, **d**, Conformation of the active site in the ATP- $\text{Mg}^{2+}$ -bound (**c**) and ADP-bound (**d**) structures.

ATP and F6P, TEM showed that wild-type (WT) PFKP particles had the dimensions and appearance of tetramers (Fig. 3b, c). In contrast, PFKP-F649L particles were the same width but half the length of WT, consistent with dimer formation along the catalytic interface (Fig. 3b, c and Extended Data Fig. 6b). We compared the PFKP-F649L particles with those induced by the inhibitor citrate, which cause PFKM to form dimers<sup>23</sup>. In a buffer containing 1 mM citrate, we saw two sizes of particles with WT PFKP: one with dimensions of tetramers and the other with dimensions of dimers along the catalytic interface (Fig. 3b, c and Extended Data Fig. 6c), further confirming dimer formation by PFKP-F649L. The catalytic activity of PFKP-F649L was reduced 98% (Extended Data Fig. 6d) compared with WT enzyme, indicating that tetramer formation is necessary for PFK1 activity.

The structures suggest that an electrostatic interaction at the tetramer interface between Arg613 of one subunit and Glu657 of the adjacent subunit (Fig. 3a) may be important for enzyme function. This salt bridge was only observed in the ATP-bound structure but not in the ADP-bound PFKP structure or dimeric rabbit PFKM structures<sup>12</sup>, suggesting that it may contribute to maintaining an active form of the mammalian tetramer. PFKP-E657A had reduced affinity of  $\sim 4.5$  mM for F6P, compared with  $\sim 0.8$  mM in WT, and an approximately twofold decrease in maximum activity (Fig. 3d, Extended Data Fig. 6a and Extended Data Table 2). Our data indicate that hydrophobic interactions are essential for the formation of tetramers while electrostatic interactions are required for optimal enzyme activity.

The structure of PFKP provides a foundation for understanding the functional effects of somatic PFK1 mutations identified in cancers. Cancer cells rely on aerobic glycolysis to provide energy and cellular building blocks required to support rapid proliferation<sup>6</sup>. PFK1 activity is increased in cancer cell lines and primary tumour tissues<sup>24</sup> and expression of PFKP is upregulated in breast<sup>25</sup> and liver<sup>26</sup> cancers. The effect of somatic mutations in PFK1 on metabolic adaptation has not been reported. We mapped the 44 reported somatic mutations

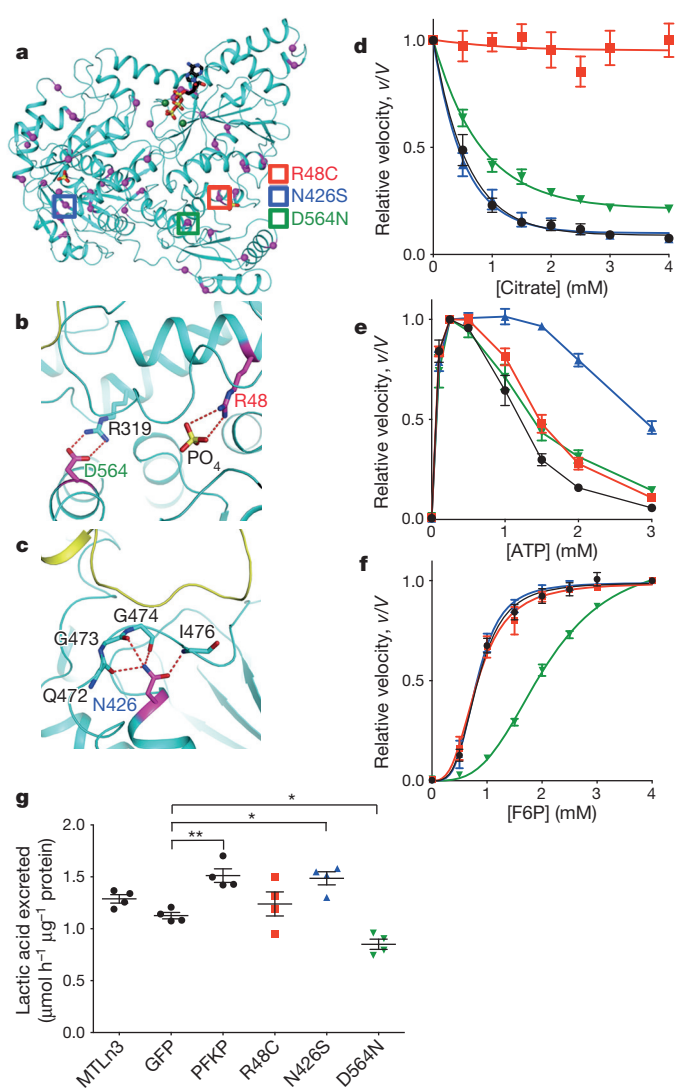




**Figure 3 | Interactions at the tetramer interface of PFKP regulate enzyme activity.** **a**, Interface of the PFKP tetramer, with one of two tetramer interfaces involving residues from subunit A (cyan) and subunit D (magenta). Two views of the hydrophobic interactions at the tetramer interface and predicted electrostatic interactions between Arg613 of one subunit with Glu657 of the adjacent subunit. Arrows indicate the position of the two-fold symmetry axis in the tetramer interface, relating the two subunits. **b**, TEM images of WT and PFKP-F649L particles in buffer containing activator and substrates or WT in buffer containing the inhibitor citrate. **c**, Scatter plot of length versus width for particles observed in TEM. WT with activator and substrate (black,  $n = 53$ ); F649L with activator and substrate (blue,  $n = 77$ ); WT with citrate, tetramers (green,  $n = 76$ ) and dimers (red,  $n = 41$ ). **d**, F6P dependence of PFKP WT (black squares) and E657A (grey circles) at 0.25 mM ATP. Data are means  $\pm$  s.e.m. of eight (WT) and five (E657A) determinations from two independent protein preparations.

in cancers<sup>27</sup> that were not associated with single nucleotide polymorphisms<sup>28</sup> onto the structure of PFKP (Fig. 4a and Extended Data Table 3). Analysis by Mutation Assessor<sup>29</sup> predicted that 28 of these mutations would alter enzyme activity.

We selected three identified somatic mutations for biochemical analysis (Extended Data Fig. 7). Arg48 interacts with a bound phosphate ion in the structure (Fig. 4b), and the R48C mutant had reduced citrate inhibition, shifting  $E_{50}^{\text{citrate}}$  from 0.4 mM for WT to greater than 4 mM (Fig. 4d) but did not markedly change effects of ATP and F6P (Fig. 4e, f and Extended Data Table 2). Analogous mutations in PFKM have been described in Tarui disease<sup>20,30</sup>. These data indicate that Arg48 is located in the citrate-binding site, which is occupied by the phosphate ion in the current structure<sup>30</sup>. A serine substitution for Asn426, located close to the catalytic interface, is predicted to disrupt interactions with the backbone carbonyls of Gln472, Gly473 and Gly474 and the main-chain amide of Ile476, which are involved in positioning a loop at the catalytic interface (Fig. 4c). The N426S mutant partly relieves ATP inhibition, shifting  $E_{50}^{\text{ATP}}$  from  $\sim 1$  mM to greater than 3 mM (Fig. 4e). Located across the catalytic interface from Asn426, Asp564 forms an electrostatic interaction with Arg319 (Fig. 4b). The D564N mutant had decreased maximum velocity and affinity for F6P (Fig. 4f and Extended Data Table 2). We also stably



**Figure 4 | Somatic cancer mutations of PFKP alter enzymatic activity and allosteric regulation.** **a–c**, Location of indicated PFKP mutations in human cancers identified from the COSMIC database and mapped onto the catalytic interface of the PFKP subunit. Mutations chosen for further analysis are denoted with coloured boxes, including Arg48Cys (a, red) and location of Arg48 at  $\text{PO}_4^{2-}$ -binding site (b), Asp564Asn (a, green) and ionic bond of Asp564 with Arg319 (b), and Asn426Ser (a, blue) and location of Asn425 at the catalytic interface (c). **d–f**, The effect of mutations on citrate inhibition (d), ATP activation and inhibition (e) and affinity for F6P (f). Data are means  $\pm$  s.e.m. of seven (d), five (e) and seven (f) determinations from two independent protein preparations. WT, black circles; Arg48Cys, red squares; Asn426Ser, blue triangles; Asn564N, green triangles. **g**, Lactic acid excretion (micromoles of lactate excreted per hour per microgram of total cell lysate) from MTLn3 rat mammary adenocarcinoma cells expressing WT and mutant PFKP-GFP. Data are means  $\pm$  s.e.m. of four experiments. \* $P < 0.05$ ; \*\* $P < 0.01$ .

expressed PFKP WT and mutants tagged with green fluorescent protein (GFP) in MTLn3 rat mammary adenocarcinoma cells (Extended Data Fig. 7b). In cell lysates, PFK1 activity was greater with expression of WT, N426S and D564N but not R48C compared with untransfected or GFP controls (Extended Data Fig. 7c). Lactic-acid excretion was also greater with cells expressing WT and N426S but significantly less with D564N compared with GFP controls (Fig. 4g). Inhibition of glycolytic flux by loss of function mutations, such as D564N, may confer a selective advantage for cancer cell growth and metastasis by redirecting carbon flow through the pentose phosphate pathway, similar to that observed by glycosylation-dependent PFK1 inhibition<sup>4</sup>. However, relief of inhibition by allosteric regulators had no effect on lactate

production in the glutamine-free cell-culture conditions we used. This finding could reflect the ability of PFK1 to dynamically alter metabolic states by integrating multiple signals. Additionally, the functional significance of selective PFKP mutations will depend on the mutational signature of the respective cancer in which they occur as well as the relative expression of other PFK1 isoforms.

In addition to cancer, aberrant glycolytic flux is increasingly recognized as contributing to several other diseases such as obesity, diabetes and Tardieu disease. The biologically relevant tetrameric structures of the PFKP provide information on the catalytic interface and conformational changes upon ATP hydrolysis that contributes to a mechanistic understanding of the functional impact of disease-associated mutations. Additionally, these new structural insights will enable rational drug design for therapeutic development.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 5 August 2014; accepted 3 March 2015.**

**Published online 18 May 2015.**

- Schöneberg, T., Kloos, M., Brüser, A., Kirchberger, J. & Sträter, N. Structure and allosteric regulation of eukaryotic 6-phosphofructokinases. *Biol. Chem.* **394**, 977–993 (2013).
- Tardieu, S. *et al.* Phosphofructokinase deficiency in skeletal muscle. A new type of glycogenosis. *Biochem. Biophys. Res. Commun.* **19**, 517–523 (1965).
- Getty-Kaushik, L. *et al.* Mice deficient in phosphofructokinase-M have greatly decreased fat stores. *Obesity* **18**, 434–440 (2010).
- Yi, W. *et al.* Phosphofructokinase 1 glycosylation regulates cell growth and metabolism. *Science* **337**, 975–980 (2012).
- Moreno-Sánchez, R. *et al.* Phosphofructokinase type 1 kinetics, isoform expression and gene polymorphisms in cancer cells. *J. Cell. Biochem.* **113**, 1692–1703 (2012).
- Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
- Dunaway, G. A. A review of animal phosphofructokinase isozymes with an emphasis on their physiological role. *Mol. Cell. Biochem.* **52**, 75–91 (1983).
- Evans, P. R., Farrants, G. W. & Hudson, P. J. Phosphofructokinase: structure and control. *Phil. Trans. R. Soc. Lond. B* **293**, 53–62 (1981).
- Rypniewski, W. R. & Evans, P. R. Crystal structure of unliganded phosphofructokinase from *Escherichia coli*. *J. Mol. Biol.* **207**, 805–821 (1989).
- Paricharttanakul, N. M. *et al.* Kinetic and structural characterization of phosphofructokinase from *Lactobacillus bulgaricus*. *Biochemistry* **44**, 15280–15286 (2005).
- Mosser, R., Reddy, M. C. M., Bruning, J. B., Sacchetti, J. C. & Reinhart, G. D. Redefining the role of the quaternary shift in *Bacillus stearothermophilus* phosphofructokinase. *Biochemistry* **52**, 5421–5429 (2013).
- Banaszak, K. *et al.* The crystal structures of eukaryotic phosphofructokinases from baker's yeast and rabbit skeletal muscle. *J. Mol. Biol.* **407**, 284–297 (2011).
- Sträter, N. *et al.* Molecular architecture and structural basis of allosteric regulation of eukaryotic phosphofructokinases. *FASEB J.* **25**, 89–98 (2011).
- Mcnae, I. W. *et al.* The crystal structure of ATP-bound phosphofructokinase from *Trypanosoma brucei* reveals conformational transitions different from those of other phosphofructokinases. *J. Mol. Biol.* **385**, 1519–1533 (2009).
- Sánchez-Martínez, C., Estévez, A. M. & Aragon, J. J. Phosphofructokinase C isozyme from ascites tumor cells: cloning, expression, and properties. *Biochem. Biophys. Res. Commun.* **271**, 635–640 (2000).
- Ferreras, C., Hernández, E. D., Martínez-Costa, O. H. & Aragón, J. J. Subunit interactions and composition of the fructose 6-phosphate catalytic site and the fructose 2,6-bisphosphate allosteric site of mammalian phosphofructokinase. **284**, 9124–9131 (2009).
- Rizzo, S. C. & Eckel, R. E. Control of glycolysis in human erythrocytes by inorganic phosphate and sulfate. *Am. J. Physiol.* **211**, 429–436 (1966).
- Akerman, J. W., Gorter, G., Sixma, J. J. & Staal, G. E. Human platelet 6-phosphofructokinase. Purification, kinetic parameters and the influence of sulphate ions on enzyme activity. *Biochim. Biophys. Acta* **370**, 102–112 (1974).
- Schirmer, T. & Evans, P. R. Structural basis of the allosteric behaviour of phosphofructokinase. *Nature* **343**, 140–145 (1990).
- Hesterberg, L. K. & Lee, J. C. Self-association of rabbit muscle phosphofructokinase: effects of ligands. *Biochemistry* **21**, 216–222 (1982).
- Leite, T., Da Silva, D., Coelho, R., Zancan, P. & Sola-Penna, M. Lactate favours the dissociation of skeletal muscle 6-phosphofructo-1-kinase tetramers down-regulating the enzyme and muscle glycolysis. *Biochem. J.* **408**, 123–130 (2007).
- Zancan, P., Marinho-Carvalho, M. M., Faber-Barata, J., Dellias, J. M. M. & Sola-Penna, M. ATP and fructose-2,6-bisphosphate regulate skeletal muscle 6-phosphofructo-1-kinase by altering its quaternary structure. *IUBMB Life* **60**, 526–533 (2008).
- Telford, J. N., Lad, P. M. & Hammes, G. G. Electron microscope study of native and crosslinked rabbit muscle phosphofructokinase. *Proc. Natl Acad. Sci. USA* **72**, 3054–3056 (1975).
- Yalcin, A., Telang, S., Clem, B. & Chesney, J. Regulation of glucose metabolism by 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatases in cancer. *Exp. Mol. Pathol.* **86**, 174–179 (2009).
- Moon, J.-S. *et al.* Krüppel-like factor 4 (KLF4) activates the transcription of the gene for the platelet isoform of phosphofructokinase (PFKP) in breast cancer. *J. Biol. Chem.* **286**, 23808–23816 (2011).
- Park, Y.-Y. *et al.* Tat-activating regulatory DNA-binding protein regulates glycolysis in hepatocellular carcinoma by regulating the platelet isoform of phosphofructokinase through microRNA 520. *Hepatology* **58**, 182–191 (2013).
- Forbes, S. A. *et al.* COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* **39**, D945–D950 (2010).
- A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
- Reva, B., Antipin, Y. & Sander, C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res.* **39**, e118 (2011).
- Li, Y., Rivera, D., Ru, W., Gunasekera, D. & Kemp, R. G. Identification of allosteric sites in rabbit phosphofructo-1-kinase. *Biochemistry* **38**, 16407–16412 (1999).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank S. Oakes for access to FPLC instrumentation, R. Fletterick and K. White for discussions, and A. Lauricella and G. DeTitta for the crystallization screening. This work was supported by National Institutes of Health R01 GM047413 to D.L.B., a Canadian Institutes of Health Research postdoctoral fellowship and a Pilot/Feasibility grant from the University of California, San Francisco, Liver Center (P30 DK026743) to B.A.W., and a grant from the Protein Structure Initiative of the National Institutes of Health U54-GM094597 to L.T.

**Author Contributions** B.A.W. and D.L.B. conceived initial studies with recombinant PFKP. B.A.W. expressed and purified recombinant PFKP, and performed thermostability screens to identify buffer conditions for protein stability. Crystal screening hits were extensively optimized by F.E.S. and F.F. J.S. performed X-ray diffraction data collection and processing. F.F. determined and refined the structures. B.A.W. generated and biochemically characterized recombinant WT and mutant PFKP, and generated and analysed cells with heterologous PFKP expression. B.A.W., F.F., L.T. and D.L.B. contributed to writing the manuscript.

**Author Information** Structures of ATP-Mg<sup>2+</sup>- and ADP-bound PFKP have been deposited in the Protein Data Bank (PDB) under accession numbers 4XYJ and 4XYK, respectively. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.L.B. (diane.barber@ucsf.edu) or L.T. (ltong@columbia.edu).

## METHODS

No statistical methods were used to predetermine sample size.

**Cloning, expression and purification of recombinant human PFKP.** *Homo sapiens* PFKP complementary DNA (cDNA) (NM\_002627.4) encoding the 784-amino-acid isoform 1 was cloned into the pFastBac HTa vector and baculovirus was generated using the Bac-to-Bac Expression system (Invitrogen) according to the manufacturer's protocols. Two billion sf21 or Hi5 cells were used to express PFKP at a multiplicity of infection of 1 for 48 h. Cell pellets were resuspended in lysis buffer (20 mM tris(hydroxymethyl)aminomethane (Tris-HCl; pH 7.5); 50 mM potassium phosphate; 1 mM 2-mercaptoethanol; 10% glycerol; 10 mM imidazole; cOmplete Protease Inhibitor Cocktail tablet (Roche)) and lysed with 15 passes of a dounce homogenizer. Cell debris was removed by centrifugation and the pellet discarded. The supernatant was incubated with Talon resin (Clontech), washed with 20 bed volumes of lysis buffer and eluted with a minimal volume of elution buffer (lysis buffer with 100 mM imidazole). Protein was concentrated using an Amicon Ultracel-30K Centrifugal Filter Unit (Millipore) and buffer exchanged into FPLC buffer (20 mM HEPES, pH 7.5, 100 mM KCl, 1 mM TCEP, 1 mM ATP, 1 mM  $\text{MgCl}_2$ , and 5% glycerol). PFKP was passed over a Superose 6 10/300 GL column (GE Healthcare) and the peak corresponding to the tetrameric fraction collected. Buffer was exchanged to crystallization buffer (20 mM HEPES, pH 7.5, 100 mM KCl, 1 mM TCEP, 10 mM  $\text{MgCl}_2$ , and 5% glycerol) containing either 10 mM ADP or 10 mM ATP using an Amicon Ultracel-30K Centrifugal Filter Unit and recombinant PFKP concentrated to  $>5 \text{ mg ml}^{-1}$ . Protein was stored at 4 °C. Recombinant PFKP was tested for activity and allosteric regulation before crystallization.

**PFK1 activity assays.** Activity assays for PFK1 were performed using an auxiliary enzyme assay<sup>31</sup>. Kinetic studies were performed in 200  $\mu\text{l}$  reaction containing 50 mM HEPES pH 7.4, 100 mM KCl, 10 mM  $\text{MgCl}_2$ , 0.15 mM NADH, 0.675 units  $\text{ml}^{-1}$  aldolase, 5 units  $\text{ml}^{-1}$  triosephosphate isomerase and 2 units  $\text{ml}^{-1}$  glycerol phosphate dehydrogenase. ATP and F6P were used as indicated. Auxiliary enzymes were de-salted using an Amicon Ultracel-10K Centrifugal Filter Unit before use. The concentration of PFKP was normalized and samples diluted as a 10 $\times$  stock in 10% glycerol, 20 mM Tris-HCl (pH 7.5) and 1 mM DTT immediately before the assay. The temperature was equilibrated to 25 °C for 10 min before initiating the reaction with the addition of PFKP. The absorbance at 340 nm was measured using a SpectraMax M5 microplate reader (Molecular Devices). Kinetic parameters were generated by linear regression analysis of the Hill equation using Prism (GraphPad Software) and are the average of a minimum of three measurements from two independent preparations of protein ( $R^2 > 0.95$  for all analyses). An unpaired *t*-test with equal variance was used to compare the activity of WT and F649L PFKP. One unit (U) of activity is defined as the amount of enzyme that catalyses the formation of 1  $\mu\text{mol}$  of fructose-1,6-bisphosphate per minute at 25 °C. Data on the effect of sulphate on PFK1 activity were obtained in the presence of 10 mM sodium sulphate or 10 mM sodium chloride as a control.

**Transmission electron microscopy.** Twenty microlitres of 25  $\mu\text{g ml}^{-1}$  PFKP was applied to glow-discharged carbon-coated grids and stained with 2% (w/v) uranyl acetate. Grids were examined and photographed with a JEOL 100CX II. For estimation of size of PFKP dimers and tetramers, the length and width of individual particles from TEM images were measured using FIJI ImageJ software<sup>32</sup>. The average length and width  $\pm$  s.d. are reported. For experiments analysing the shape and size of PFKP for crystallography studies, the protein was diluted in TEM buffer (20 mM HEPES, pH 7.5, 100 mM KCl, 1 mM DTT, 1 mM ATP, 1 mM  $\text{MgCl}_2$  and 5% glycerol). For experiments analysing the oligomeric state of the enzyme in the presence of activators, WT and F649L PFKP were diluted in TEM buffer containing 3 mM ADP, 3 mM ATP and 8 mM F6P. For experiments analysing the oligomeric state of the enzyme in the presence of inhibitors, PFKP was diluted in TEM buffer containing 1 mM citrate.

**Crystallization and structure determination.** PFKP was crystallized in two different complexes of ADP and  $\text{ATP-Mg}^{2+}$  by a microbatch method at 18 °C. For the ADP complex, 2  $\mu\text{l}$  of protein solution containing PFKP (6.35  $\text{mg ml}^{-1}$ ) was mixed with 1  $\mu\text{l}$  of the precipitant solution consisting of 200 mM potassium sodium tartrate tetrahydrate, pH 7.4, and 20% (w/v) PEG 3350. The same protein buffer was used for growing the crystals of the PFKP in complex with ATP and  $\text{Mg}^{2+}$ . The crystals were obtained using microbatch method and the precipitant solution comprising 200 mM potassium thiocyanate, pH 7, and 20% (w/v) PEG 3350. All crystals were cryoprotected by addition of 20% (v/v) ethylene glycol in the respective mother liquor and flash-frozen in liquid nitrogen for data collection at 100 K.

Crystals of the PFKP complexes both belong to space group  $P2_1$ . However, the crystallographic asymmetric unit of the ADP-bound form contains four subunits of PFKP that are assembled as one tetramer, whereas that of  $\text{ATP-Mg}^{2+}$  form

contains two tetramers. A single-wavelength native data set to resolution 3.1 Å was collected at the X4C beamline of the National Synchrotron Light Source. The diffraction images were processed with the HKL package<sup>33</sup>. The structure of PFK from rabbit skeletal muscle (PDB accession number 3O8L)<sup>12</sup> was used to determine the  $\text{ATP-Mg}^{2+}$  structure of PFKP using the molecular replacement method, with the program MolRep<sup>34</sup>. Only a monomeric model of PFK from rabbit skeletal muscle resulted in a solution, which led to structure determination of the ATP-bound PFKP structure. The remainder of the PFKP model was built manually with the program XtalView<sup>35</sup>. The structure refinement was performed with CNS<sup>36</sup>. A similar methodology was used for data collection and processing of the ADP-bound structure of PFKP, the crystal of which diffracted to 3.4 Å at the X4C beamline of the National Synchrotron Light Source. The ADP-bound structure was subsequently determined using a monomeric model of the ATP-bound complex of PFKP, with the program MolRep<sup>34</sup> followed by structure refinement by CNS<sup>36</sup>. The data processing and refinement statistics are summarized in Extended Data Table 1. The Ramachandran plots suggest that 88.1% and 81.9% of residues in ATP-bound complex and ADP-bound complex of PFKP are in most favoured regions, and there is no residue in disallowed regions, respectively. The trajectory between the  $\text{ATP-Mg}^{2+}$ -bound and the ADP-bound structures was generated using UCSF Chimera<sup>37</sup>. The structures were aligned with Matchmaker tool and the trajectory calculated with the Morph Conformation tool.

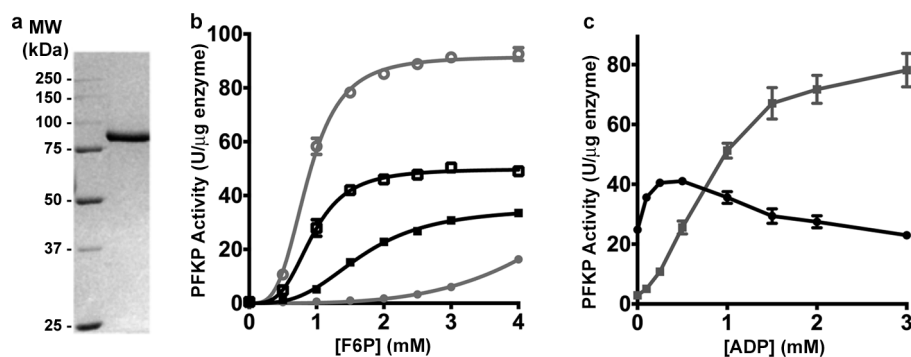
**Selection of cancer mutations and generation of point mutants.** Somatic mutations identified in human cancers were selected from the COSMIC database<sup>27</sup> and known single nucleotide polymorphisms were disregarded<sup>28</sup>. The mutations were modelled onto the structure of PFKP and selected for further analysis. Point mutants at the tetramer interface, F649L and E657A, and cancer mutants, R48C N426S, and D564N, were generated by using a commercially available site-directed mutagenesis kit (QuikChange Lightning, Agilent). DNA primers were designed using the online primer design tool (<http://www.genomics.agilent.com/primerDesignProgram.jsp>) and purchased from Elim Biopharmaceuticals.

**Analysis of cells expressing PFKP.** A mammalian PFKP expression construct was generated by PCR amplification and the cDNA inserted into the multiple cloning site of pEGFP-N1 using the restriction enzymes XhoI and BamHI. Cancer mutations were generated by site-directed mutagenesis as described above. Constructs were expressed by transfecting MTLn3 rat mammary adenocarcinoma cells<sup>38</sup> using FugeneHD (Promega) transfection reagent. One-day post-transfection cells were re-plated into 100 mm dishes and 800  $\mu\text{g ml}^{-1}$  G418 was added to select for transfected cells. After 1 week of selection, fluorescence-activated cell sorting was used to sort cells expressing GFP. For metabolic assays, cells were seeded into a six-well plate at a density of  $3 \times 10^5$  cells per well. One day after re-plating, cells were washed twice in serum- and glutamine-free media, and cells were incubated for 2 h in 1 ml of the same media. Fifty microlitres of the media were collected in triplicate and the amount of lactic acid in the media measured using an enzyme-linked assay<sup>39</sup>. One hundred microlitres of reagent A (300 mM hydrazine; 200 mM glycine, pH 9.5; 20 mM  $\beta$ -nicotinamide adenine dinucleotide) and 50  $\mu\text{l}$  of reagent B (200 U  $\text{ml}^{-1}$  L-lactate dehydrogenase from rabbit muscle (Sigma Aldrich)) were added to each well and incubated for 1 h at 22 °C. The absorbance at 340 nm was measured and the amount of lactate was determined from a standard curve. Cells were lysed in buffer (10 mM potassium phosphate, pH 7.5; 0.1% Triton X-100; cOmplete Protease Inhibitor Cocktail tablet (Roche)), cellular debris removed by centrifugation and the protein concentration determined by the Bradford method. Lactic acid levels in the media were normalized to protein concentration. PFK1 activity assays were performed on the lysate as previously described<sup>40</sup>. Enzyme-linked PFK1 activity assays were performed on 10  $\mu\text{g}$  of total cell lysate as described above with the exception that 10 mM ammonium sulphate was added to the assay mixture and the auxiliary enzymes were not de-salted. Levels of PFKP expression were determined by immunoblotting using rabbit anti-GFP (Invitrogen A-11122, 1DB-001-0000868907) and mouse anti-actin clone C4 (ED Millipore MAB1501, 1DB-001-0000850281) antibodies. Two-sided paired *t*-tests were used to determine statistical significance.

- Brüser, A., Kirchberger, J., Kloos, M., Sträter, N. & Schöneberg, T. Functional linkage of adenine nucleotide binding sites in mammalian muscle 6-phosphofructokinase. *J. Biol. Chem.* **287**, 17546–17553 (2012).
- Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nature Methods* **9**, 676–682 (2012).
- Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- Vagin, A. & Teplov, A. Molecular replacement with MOLREP. *Acta Crystallogr.* **66**, 22–25 (2010).
- McRee, D. E. XtalView/Xfit—a versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125**, 156–165 (1999).
- Brünger, A. T. et al. Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).



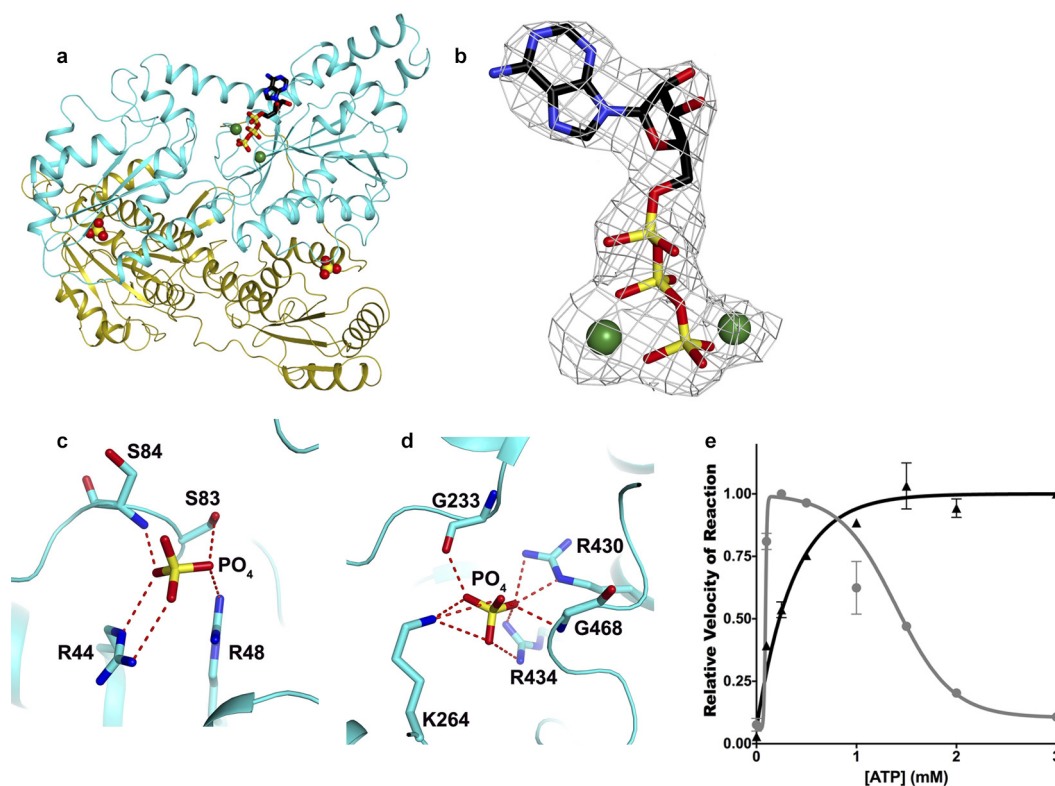
37. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
38. Segall, J. E. *et al.* EGF stimulates lamellipod extension in metastatic mammary adenocarcinoma cells by an actin-dependent mechanism. *Clin. Exp. Metastasis* **14**, 61–72 (1996).
39. Lundholm, L., Mohme-Lundholm, E. & Vamos, N. Lactic acid assay with L(+)lactic acid dehydrogenase from rabbit muscle. *Acta Physiol. Scand.* **58**, 243–249 (1963).
40. Abrantes, J. L. *et al.* Herpes simplex type 1 activates glycolysis through engagement of the enzyme 6-phosphofructo-1-kinase (PFK-1). *Biochim. Biophys. Acta Mol. Basis Dis.* **1822**, 1198–1206 (2012).



### Extended Data Figure 1 | Activity of purified recombinant PFKP.

**a**, Coomassie-stained SDS–polyacrylamide gel electrophoresis (SDS–PAGE) of purified PFKP. The molecular mass (MM) of protein standards is shown in kilodaltons (kDa). **b**, Allosteric regulation of PFKP by ATP and ADP; F6P saturation curve PFKP in the presence of 0.25 mM ATP (filled black squares),

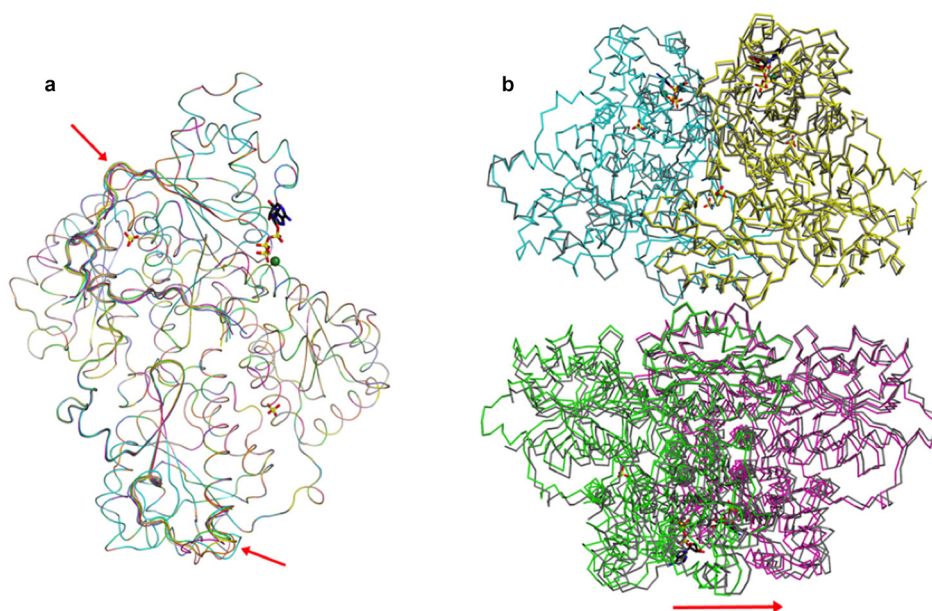
3 mM ATP (filled grey circles), 0.25 mM ATP and 0.25 mM ADP (open black squares), and 3 mM ATP and 3 mM ADP (open grey circles). **c**, Effect of ADP on kinetic behaviour of PFKP in the presence of 0.25 mM (black squares) or 3 mM (grey circles) ATP. Data in **b** and **c** are means  $\pm$  s.e.m. of ten (**b**) or five (**c**) determinations from two separate protein preparations.



**Extended Data Figure 2 | Structure, nucleotide binding and phosphate ion binding of PFKP.** **a**, The structure of PFKP protomer can be divided into two halves: the amino (N)-terminal (cyan) and the carboxy (C)-terminal (yellow) subdomains. The N terminus of each subdomain begins with a nucleotide-binding domain (NBD) followed by a smaller substrate-binding domain (SBD). Each NBD closely resembles a canonical Rossmann fold composed of a seven-stranded  $\beta$ -sheet surrounded by six  $\alpha$ -helices. Each SBD consists of a four-stranded  $\beta$ -sheet surrounded by five  $\alpha$ -helices. Two phosphate ions (stick drawings) are bound in pockets equivalent to the effector binding sites of the *E. coli* PFK. **b**, Final  $2F_o - F_c$  electron density at 3.1 Å resolution for ATP-Mg<sup>2+</sup>, contoured at  $1\sigma$ . A strong electron density is observed near  $\beta$ - and

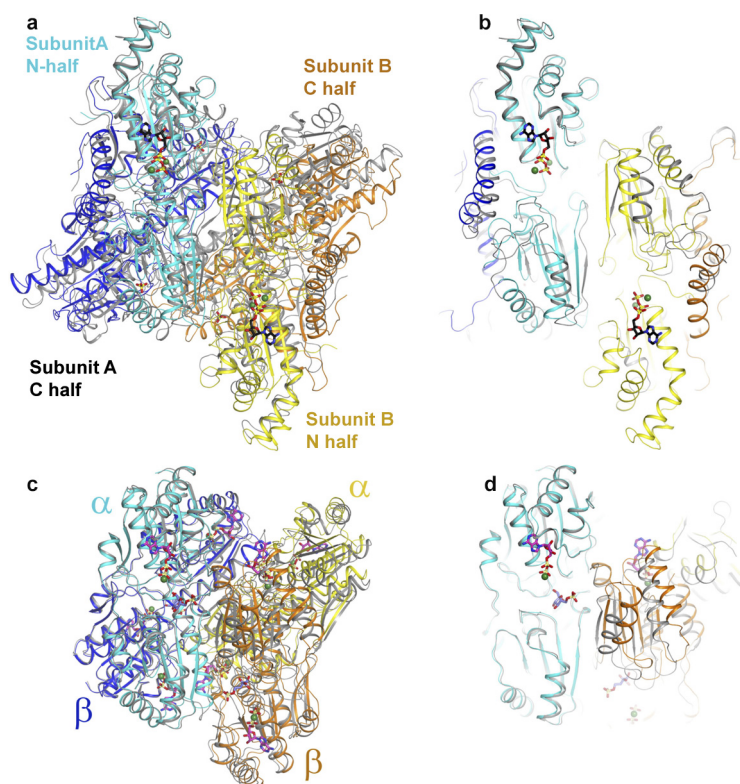
$\gamma$ -phosphate of the nucleotide, which was unambiguously modelled as Mg<sup>2+</sup> ion. An extended but weaker electron density is also observed near the  $\gamma$ -phosphate of the nucleotide, which is surrounded by three backbone carbonyls of strictly conserved Ser32, Gly34 and Gly172. This electron density was modelled as a second metal ion, although it may belong to a water molecule. **c, d**, Structure of the two inorganic phosphate-binding sites in PFKP. **e**, Plot of concentration of ATP versus relative enzymatic activity of PFKP in the presence (black triangles) and absence (grey circles) of 10 mM sodium sulphate. Activity is expressed relative to maximal activity at this pH and F6P concentration. Data are means  $\pm$  s.e.m. of three determinations.





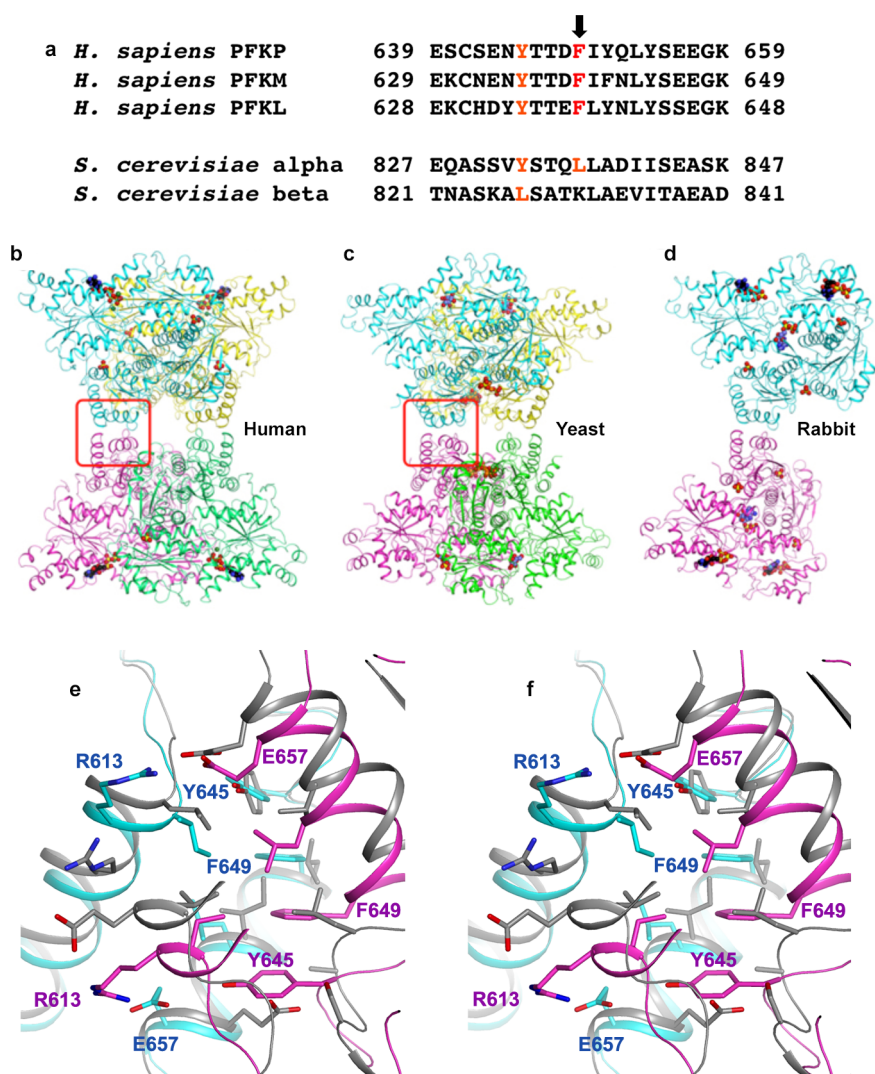
**Extended Data Figure 3 | Structural comparison of the two PFKP tetramers in the ATP-Mg<sup>2+</sup> complex.** **a**, Overlay of the structures of the eight PFKP subunits. Only two loops show substantial differences, indicated with the red

arrows. **b**, Overlay of the two PFKP tetramers. A noticeable difference is the twisting of the second dimer in the two tetramers, indicated with the red arrow.



**Extended Data Figure 4 | Structural comparison of ATP- and ADP-bound PFKP with R- and T-state of *E. coli* PFK.** **a**, Structural overlay of ATP-bound (coloured) and ADP-bound (grey) PFKP. For comparison with structures of *E. coli* PFK, the N- and C-terminal subdomains of PFKP are coloured cyan and

blue for subunit A, and yellow and orange for subunit B. **b**, The view in **a** is slabbbed to highlight the difference between the two structures. **c**, Structural overlay of R-state (coloured; PDB accession number 4PFK)<sup>11</sup> and T-state (grey; PDB accession number 6PFK)<sup>23</sup> of *E. coli* PFK. **d**, The view in **c** is slabbbed.



#### Extended Data Figure 5 | A unique tetramer interface in PFKP.

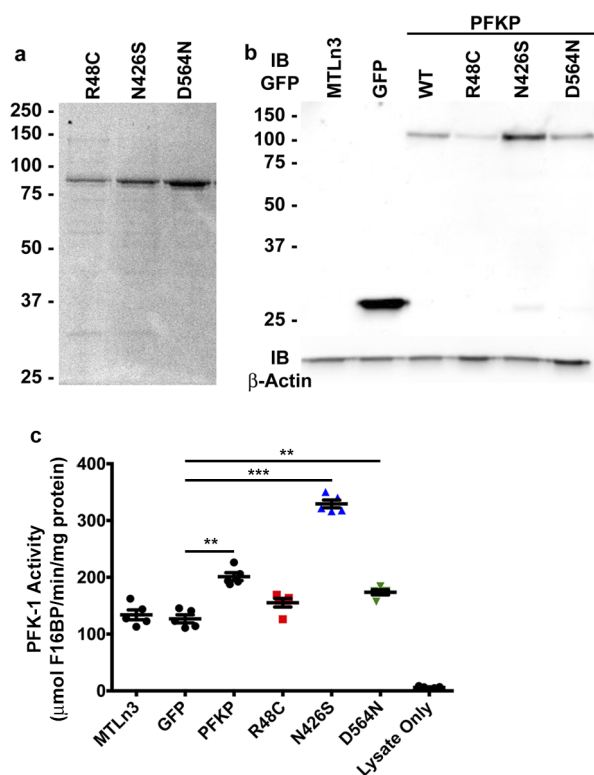
**a**, Alignment of residues from PFKP surrounding Phe649 (arrow) with human PFKM and PFKL and *S. cerevisiae* PFK1  $\alpha$ - and  $\beta$ -subunits. **b**, Structure of PFKP tetramer. **c**, Structure of ScPFK tetramer<sup>15</sup>, viewed roughly in the same

orientation as PFKP. The tetramer interface is highlighted in the red box.

**d**, Structure of rabbit PFKM<sup>15</sup>. **e**, Stereo drawing of the overlay of the tetramer interface of PFKP (in colour) and ScPFK.







**Extended Data Figure 7 | Purification of PFKP cancer mutants and their activity in cells.** **a**, Coomassie-stained SDS-PAGE of purified recombinant PFKP mutants R48C, N426S and D564N. **b**, Immunoblot of GFP and actin from total cell lysates of MTLn3 rat mammary adenocarcinoma cells expressing PFKP-GFP. Blots are representative of three experiments from individual preparations of cells. **c**, PFK1 activity (micromoles of F1,6bP produced per minute per nanogram of total cell lysate) was measured in five independent preparations of cells. A two-sided paired *t*-test was used to determine significance. \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ .

Extended Data Table 1 | Data collection and refinement statistics

|   | PFKP<br>(ATP-Mg <sup>2+</sup> complex) | PFKP<br>(ADP complex)   |
|---|--|-------------------------|
| <b>Data collection</b>                              |  |                         |
| Space group   | <i>P</i> 2 <sub>1</sub>                | <i>P</i> 2 <sub>1</sub> |
| Cell dimensions                                     |  |                         |
| <i>a</i> , <i>b</i> , <i>c</i> (Å)                  | 137.2, 159.3, 170.5                    | 79.3, 168.4, 133.3      |
| $\alpha$ , $\beta$ , $\gamma$ (°)                   | 90, 104.2, 90                          | 90, 103.8, 90           |
| Resolution (Å)                                      | 50-3.1 (3.2-3.1) *                     | 45.4-3.4 (3.5-3.4) *    |
| <i>R</i> <sub>merge</sub>                           | 11.7 (68.4)                            | 15.6 (53.4)             |
| <i>I</i> / $\sigma$ <i>I</i>                        | 13.9 (1.8)                             | 4.3 (1.2)               |
| Completeness (%)                                    | 93.0 (84.1)                            | 84.8 (74.3)             |
| Redundancy  | 6.4 (5.3)                              | 2.1 (1.9)               |
| <b>Refinement</b>                                   |  |                         |
| Resolution (Å)                                      | 50-3.1 (3.3-3.1)                       | 45.4-3.4 (3.6-3.4)      |
| No. reflections                                     | 109,577 (10,980)                       | 34,657 (3,486)          |
| <i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub> | 22.8/25.8                              | 24.2/29.4               |
| No. atoms   | 47,148                                 | 23,644                  |
| Protein   | 46,792                                 | 23,500                  |
| Ligand/ion  | 316                                    | 144                     |
| Water   | 40                                     | 0                       |
| B-factors   |  |                         |
| Protein   | 62.8                                   | 74.5                    |
| Ligand/ion  | 45.6                                   | 66.1                    |
| Water   | 22.3                                   |                         |
| R.m.s deviations                                    |  |                         |
| Bond lengths (Å)                                    | 0.009                                  | 0.011                   |
| Bond angles (°)                                     | 1.3                                    | 1.2                     |

One crystal was used for data collection for each structure.

\* Highest resolution shell is shown in parenthesis.



**Extended Data Table 2 | Saturation kinetics on WT and mutant PFKP**

| Parameter                | Wild type | E657A | R48C  | N426S | D564N |
|--------------------------|-----------|-------|-------|-------|-------|
| Maximum Velocity         | 59.27     | 32.29 | 58.19 | 67.41 | 30.60 |
| $S_{0.5}^{F6P}$ (mM)     | 0.83      | 4.51  | 0.84  | 0.82  | 2.04  |
| $n_H^{F6P}$ (mM)         | 3.41      | 4.13  | 2.94  | 3.64  | 3.18  |
| $EC_{50}^{ATP}$ (mM)     | 0.96      | ND    | 1.19  | >3    | 0.68  |
| $EC_{50}^{Citrate}$ (mM) | 0.40      | ND    | >4    | 0.31  | 1.40  |

Kinetic properties of WT and mutant protein were determined by modelling the sigmoidal part of the curve ( $V_{min}$  to  $V_{max}$ ) to the Hill equations. Assays were performed at pH 7.4 with 0.25 mM ATP. For F6P affinity, assays were performed at pH 7.4 with 0.25 mM ATP. ATP and citrate inhibition assays were performed at pH 7.4 with 2 mM F6P (WT, R48C and N426S) or 4 mM F6P (D564N). Citrate inhibition assays were performed with 0.25 mM ATP. ND, not determined.

Extended Data Table 3 | Somatic mutations of PFKF in cancer

| Missense Mutation* | Mutation ID* | Ligand Interactions** | Predicted Impact on Activity** | Tissue/Cancer Type*                                |
|--------------------|--------------|-----------------------|--------------------------------|--|
| S32R               | 1603385      | ADP ATP               | High                           | Liver Carcinoma                                    |
| R48C               | 1347553      | PGA ADP               | High                           | Large Intestine Carcinoma                          |
| R48H               | 917671       | PGA ADP               | Medium                         | Endometrium Carcinoma                              |
| M49L               | 1675020      |                       | Medium                         | Lung Carcinoma                                     |
| I51V               | 538534       |                       | Medium                         | Lung Carcinoma                                     |
| D128G              | 1492251      | F1,6bP ADP ATP        | High                           | Kidney Carcinoma                                   |
| G129W              | 1187879      | ADP ATP               | High                           | Lung Carcinoma                                     |
| L131I              | 330744       | ATP                   | High                           | Lung Carcinoma                                     |
| Q153K              | 1236264      |                       | Neutral                        | Autonomic ganglia Neuroblastoma                    |
| A158V              | 917677       |                       | Low                            | Endometrium Carcinoma                              |
| D175Y              | 1347563      | ATP F6P F1,6bP        | High                           | Large Intestine Carcinoma                          |
| R219Q              | 1603390      | ATP ADP F6P F1,6bP    | High                           | Liver Carcinoma                                    |
| E245Q              | 684504       | ATP ADP               | Low                            | Lung Carcinoma                                     |
| R262Q              | 1347567      | PGA ADP               | Low                            | Large Intestine Carcinoma                          |
| E286K              | 1474583      |                       | Medium                         | Breast Carcinoma                                   |
| V293I              | 255484       |                       | Medium                         | Primitive neuroectodermal tumour - medulloblastoma |
| R301H              | 1347571      | F6P                   | High                           | Large Intestine Carcinoma                          |
| V308M              | 917679       |                       | Medium                         | Endometrium Carcinoma                              |
| E328V              | 1603392      |                       | Medium                         | Liver Carcinoma                                    |
| A332T              | 291604       |                       | Medium                         | Large Intestine Carcinoma                          |
| P407S              | 233090       |                       | Neutral                        | Skin Malignant melanoma                            |
| A414D              | 1347575      |                       | High                           | Large Intestine Carcinoma                          |
| N426S              | 917681       |                       | High                           | Endometrium Carcinoma                              |
| A445T              | 1220252      |                       | Medium                         | Large Intestine Carcinoma                          |
| W463C              | 370998       | PGA ADP               | High                           | Lung Carcinoma                                     |
| G467A              | 332540       | PGA ADP               | Neutral                        | Lung Carcinoma                                     |
| T470I              | 1702000      |                       | Neutral                        | Skin Malignant melanoma                            |
| A492T              | 1675022      | F1,6bP                | High                           | Lymphoid neoplasm                                  |
| A537S              | 26929        |                       | Low                            | Lung Carcinoma                                     |
| D564N              | 241118       |                       | Medium                         | Large Intestine Carcinoma                          |
| R575Q              | 1220254      |                       | Low                            | Large Intestine Carcinoma                          |
| A603T              | 1492249      |                       | Medium                         | Kidney Carcinoma                                   |
| K627E              | 1347584      | PGA ADP               | Low                            | Large Intestine Carcinoma                          |
| K627N              | 72151        | PGA ADP               | Medium                         | Ovary Carcinoma                                    |
| D648N              | 1560809      |                       | Low                            | Large Intestine Carcinoma                          |
| N667Y              | 1347586      |                       | Medium                         | Large Intestine Carcinoma                          |
| P680A              | 917689       |                       | Low                            | Endometrium Carcinoma                              |
| I689M              | 1128065      |                       | Neutral                        | Prostate Carcinoma                                 |
| E703D              | 684499       |                       | Medium                         | Lung Carcinoma                                     |
| K709I              | 255302       |                       | Low                            | Primitive neuroectodermal tumour - medulloblastoma |
| T713A              | 1239809      |                       | Low                            | Oesophagus Carcinoma                               |
| E734G              | 98079        |                       | Low                            | Upper aerodigestive tract Carcinoma                |
| M758I              | 117577       |                       | Medium                         | Ovary Carcinoma                                    |
| L761P              | 538526       |                       | Medium                         | Lung Carcinoma                                     |

Residues highlighted in grey were chosen for further characterization.

\* Missense mutations identified, mutation identification number and the tissue/cancer type each mutation was identified from COSMIC database<sup>31</sup>.\*\* Ligand interactions and predicted impact on activity obtained from Mutation Assessor<sup>32</sup>. Mutations with 'High' or 'Medium' impact are predicted to alter enzyme activity.

## TOOLBOX

# COMPUTERS READ THE FOSSIL RECORD

*Palaeontologists hope that software can construct fossil databases directly from research papers.*

ILLUSTRATION BY THE PROJECT TWINS



BY EWEN CALLAWAY

For a field whose *raison d'être* is to chronicle the deep past, palaeontology is remarkably forward-looking when it comes to organizing its data. Victorian natural history museums meticulously organized their collections with handwritten cards that survive to this day. And over the past 15 years, researchers have collectively entered records of more than a million fossils into an online database, allowing

them to track broad trends in the history of life. Now, palaeontologists are exploring the use of machine algorithms to pull fossil data from their research papers automatically.

"I'm fairly convinced that this is the future, for sure," says Shanan Peters, a palaeontologist at the University of Wisconsin–Madison (UW Madison) who is co-leading an effort to use software to extract information from tens of thousands of palaeontology papers. "Building a database, *per se*, will be a thing of the past.

Those databases will be dynamically generated based on the questions you're interested in, and the machine will do the heavy lifting."

Peters should know. He is the principal investigator of the Paleobiology Database (PBDB; [paleobiodb.org](http://paleobiodb.org)), which details the age, location and identity of some 1.2 million fossils. Since it was started in 1998, researchers have spent about 80,000 hours — the equivalent of 9 continuous years — entering and opening over data from original field research and ►



► around 40,000 articles. The PBDB has produced hundreds of papers and has allowed palaeontologists to address questions that would have been otherwise unanswerable, on topics ranging from epoch-wide extinction rates to the disappearance of certain dinosaurs.

The PBDB is a database created by experts: around 380 scientists have uploaded some 560,000 published opinions on the classifications of 320,000 taxonomic names. But Peters was curious to know whether such a database could be compiled automatically by computer. So in 2013 he started a collaboration with Miron Livny and Chris Ré, then data scientists at UW Madison (Ré has since moved on to Stanford University in California). Ré had developed software called DeepDive, which mines written text (such as words in a research paper) and pulls out facts. Text mining — or content mining — is now a commonplace tool in computer science and is slowly beginning to find uses in research fields from genomics to drug discovery. Text mining palaeontology literature appealed to Ré, partly because the PBDB offers a human-curated database with which to compare a computer-generated counterpart.

## PARSING THE PAST

DeepDive begins by parsing research papers in a manner that would be familiar to anyone who remembers their early grammar lessons. “It’s taking those papers and converting them into text,” says Ré: it is trying to determine the answer to questions such as, “What’s a noun, what’s a verb and how do you diagram a sentence?” Next, DeepDive attempts to predict the concepts that are stored in those sentences (such as, for palaeontology, the names of fossils and the places where they were found) and assigns a probability to each assertion. The result is software “which is usually imperfect in a lot of ways”, says Ré. “That’s where you get the domain scientist involved.”

Peters spent about a year refining the first-pass software so that, for instance, it knows where to look in palaeontology papers for the names of new species and the geographic locations in which they were discovered. Ré describes this process as a “back and forth” with Peters that required Ré’s team of data scientists to come up with custom computing solutions to make the requests feasible. “I would love to say the answer is people can press a button and use it and run it and they don’t need us,” Ré says — but that is a goal that his team has not yet reached.

As a proof of principle, Peters and Ré used custom software that they called PaleoDeepDive to create a text-mined, scaled-down version of the PBDB that incorporated around 12,000 papers. In some ways the computer-generated database outshines the PBDB, Peters says, because all the information in it comes with a probability assigned to it and is linked back to the original text. “The machine is really clear about uncertainty, when there’s ambiguity, or differences between documents and authors,”

Peters says. PaleoDeepDive also managed to extract 192,000 opinions on the classification of taxonomic names from the papers, whereas the PBDB’s human curators found only 80,000.

PaleoDeepDive did not do such a bad job at organizing that information either. In a December 2014 paper, Ré and Peters report that from a random sample of 100 statements drawn from the computer-generated database, 92% were correct — which they say was similar to the accuracy of the PBDB (S. E. Peters *et al.* *PLoS ONE* **9**, e113523; 2014). The two databases also scored similarly in a second experiment, when scientists were presented with five documents and asked to score the accuracy of facts that had been mined from them by the PBDB and by the computer.

And perhaps most impressively, PaleoDeepDive was used to estimate species diversity and extinction rates over the past 500 million years, coming up with measures similar to those determined by the PBDB.

“It’s a little scary, the machines are getting that good. That’s just something that we’re going to have to get used to,” says Mark Uhen, a palaeontologist at George Mason University in Fairfax, Virginia, who is on the PBDB’s executive council. “I think it’s one of the best innovations that palaeontology has had in a very long time,” says Jonathan Tennant, a palaeontologist at Imperial College London. He uses the PBDB every day and thinks that text mining could serve as a useful way to collect a large amount of data for later manual inspection — but not as a full-on replacement for human-curated databases such as the PBDB. “I don’t see machines replacing humans. I think it’s important that we retain the human aspect of the analytics,” he says.

John Alroy, a palaeontologist at Macquarie University in Sydney, Australia, who co-founded the PBDB but is no longer affiliated with it, is less bullish on text mining. He says that DeepDive tends to overestimate the period during which species existed, leading to mistaken estimates of species diversity. He sees speed as the only advantage of text mining. “But there is no need to be fast in this case because the PBDB is already extremely comprehensive, so pretty much any question you might want to ask can already be answered with it. That explains why it has generated so many publications,” Alroy says.

## TEXT-MINING FRUSTRATIONS

Peters says that he will be using the computer-generated database as a supplement to the human-generated PBDB but adds that, for now, the limited number of documents it works from make it of little added use to palaeontologists. He wanted to let PaleoDeepDive loose on a bigger set of documents, but he did not have legal permission. As other text miners have

discovered, many publishers of paywalled articles are cautious about allowing researchers to text mine their papers, even if they have lawful access to the literature; publishers tend to place limits on how the results of text mining can be published and reused, and often limit the number of papers a scientist can download at any one time (see *Nature* **483**, 134–135; 2012). “I can’t think of any single palaeontologist who has 40,000 papers in their own stash, at least legally acquired,” says Tennant.

Peters and Livny spent months brokering a deal with one scientific publisher, Elsevier, to gain access to thousands of papers. “This is just the frustrating reality of things right now: advanced capabilities in machine reading and learning are coming out, and the bottleneck in progress is now getting documents together in one place for analysis,” Peters says. He and his colleagues are working on amassing and parsing documents to feed into PaleoDeepDive and a related software tool for the geosciences literature called GeoDeepDive. Ré, meanwhile, is working with experts in other fields to apply DeepDive to drug development, genomics and human trafficking.

Many palaeontologists also want to make it easier to find the data buried in their papers, so they are calling for research papers to be described more systematically in the future. “If we start having publication where everything is standard, then it will be much easier to read and process that data,” says Tennant. Uhen adds, “I think there’s a sort of cultural shift going on in palaeontology, where people are interested in data aggregation, and getting more insistent about being crystal clear about where you’re finding your fossils.”

Despite these challenges, many palaeontologists see text mining as the way forward for their field. “It’s a huge waste of time for grad students and postdocs to manually re-enter already published information into a structured database,” says Ross Mounce, a palaeontologist at the Natural History Museum in London who is using text mining to track how the museum’s 80-million-specimen collection is used in research papers. Peters hopes that efforts such as PaleoDeepDive will allow him and his colleagues more time to generate data instead of spending their days organizing data they already have. “I see these machine reading systems as liberating our efforts a little bit, and shifting our work back into the field and back into the museums.” ■

Ewen Callaway writes for *Nature* from London.

## CORRECTION

The Toolbox article ‘How to catch a cloud’ (*Nature* **522**, 115–116; 2015) gave the wrong location for the Texas Advanced Computing Center — it is in Austin not San Antonio.

# CAREERS

**TRADE TALK** How a science educator moved from research lab to teaching lab **p.119**

**SCIENCE COMMUNICATION** Opportunities for public engagement [go.nature.com/zex87n](http://go.nature.com/zex87n)

**NATUREJOBS** For the latest career listings and advice [www.naturejobs.com](http://www.naturejobs.com)

CLAIRE WELSH/NATURE



## LEISURE ACTIVITIES

# The power of a pastime

*From painting to punching to aeroplane-jumping, the hobbies that scientists pursue offer a vital escape from the laborious life of the lab.*

BY CHRIS WOOLSTON

Albert Einstein mastered the violin. Richard Feynman banged bongos. Following in the tradition of multi-talented physicists, Federica Bianco likes to take a break from her research to punch people in the face. Bianco, an avid boxer who is also an astrophysicist at New York University, flew to Richmond, California, for her first professional bout in April. It did not go well for her opponent. Bianco pinned her competitor to the ropes with a flurry of punches and did not let up until the referee called the fight. It took just one minute and twenty seconds. “I didn’t want to stop, but she was taking too much punishment,” Bianco says.

For Bianco, boxing is not just a hobby; it is a total mind-and-body escape from her work. “As a scientist, I’m thinking about all sorts of things all the time,” she says. “The ring is quiet. You get tunnel vision. The other person is trying to take off your head and you have to deal with that.”

At a time when competition for science funding and job promotions sometimes resembles a boxing match, many researchers have trouble conceiving of an active life outside the lab. Indeed, there can be subtle — or not so subtle — pressures to sacrifice leisure time and put aside other interests for the sake of the next experiment, paper or conference talk. But many scientists say that their pastimes make them better researchers by sharpening

their minds, building confidence and reducing stress. Their experiences should offer hope to researchers who are feeling overwhelmed by the pressure of their jobs. Release can be just a ride, jump, joke or punch away.

To be sure, some senior researchers in academia and other sectors still look askance at hobbies or leisure activities. Ryan Raver, now a product manager at the biotechnology firm Sigma-Aldrich in St Louis, Missouri, recalls an instance at the University of Wisconsin–Madison when one of his thesis-committee members was reluctant to sign off on his PhD because he thought that Raver spent too much time blogging and playing lead guitar in a hard-rock band. “He thought I should ►

► have been more focused on my work,” Raver says. “But playing in the band helped me survive grad school. It kept my excitement and motivation up. It pushed me through the day.”

Sean Carroll, a physicist at the California Institute of Technology in Pasadena, wrote a blogpost advising scientists to choose their hobbies carefully, especially if they ever want to win tenure. Specifically, he counsels them to stay away from pastimes that could drain attention from science. “You are better off if your hobbies are nothing like your work,” he writes. “Permissible hobbies include skydiving, playing guitar, or cooking. Suspicious hobbies include writing of any sort (novels, magazine articles, blogs), programming or web stuff, starting a business, etc. Why? Because there’s a feeling that this kind of activity represents time that could be spent on research.”

Carroll, who in 2005 was denied tenure at a different institution, says in his blogpost that he regrets the time and effort that he put into writing *Spacetime and Geometry: An Introduction to General Relativity* (Addison-Wesley, 2003), a textbook that did not win him many points in scientific circles. In the blogpost, he deems the book “probably the worst thing I did personally”.

Raver says that he had a much easier time following his outside passions once he left academia for an industry job. “Professors and academics want you to believe that the more

hours you put in, the more likely it is that you’ll have quality data,” he says. “But people aren’t machines. They need to take breaks and reset their minds when things get tough.”

### FIGHT SCIENCE WITH SCIENCE

There is plenty of evidence that scientific research and leisure pursuits can coexist. A study published in 2008 found that Nobel prizewinners were more likely than other scientists or members of the public to have long-standing hobbies. Notably, the prizewinners were about 1.5 times more likely to actively pursue arts and crafts than were members of the US National Academy of Sciences (R. Root-Bernstein *et al. J. Psychol. Sci. Technol.* 1, 51–63; 2008). For this sample, hobbies turned out to be better predictors of Nobel-level greatness even than reported IQ, which does not vary much between ‘top’ and ‘average’ scientists.

Robert Root-Bernstein, a physiologist at Michigan State University in East Lansing and lead author of the study, says that it is hard to know whether hobbies themselves help to fuel genius or whether geniuses are simply more likely to take up hobbies (see “The secret lives of polymaths”). “It’s probably some combination,” he says. He also notes that, contrary to public opinion, scientific masterminds tend to be more adventurous, daring and physically vigorous than are members of the general

public. “An unexpectedly large number of Nobel laureates took up surfing when it came into fashion in the 1960s,” he says.

Some in academia do appreciate the value of hobbies and leisure pursuits for early-career researchers, perhaps because they have discovered it for themselves. Tony Ryan, a chemist and pro-vice-chancellor of the fac-

**“Permissible hobbies include skydiving, playing guitar, or cooking. Suspicious hobbies include writing of any sort.”**

ulty of science at the University of Sheffield, UK, has hired many scientists over the years. He has always been reluctant to offer a position to anyone who is so focused on research that she or he has no time for anything else.

“We want to know that you’re an A-1 excellent scientist, but we also want to know that you’re a well-rounded person whom students will relate to,” he explains.

He has his own obsession: despite the demands of his job, he logs about 8,000 kilometres on his bicycle every year. He bikes to and from work, and rides with a group of enthusiasts — who call themselves ‘Common Lane Occasionals’ — every Saturday morning. While pedalling, he likes to talk about prime numbers with a computer scientist, sometimes to the annoyance of other members of the group, which includes a tree surgeon, a plumber and a physician.

Ryan even brings his bike along to international conferences, whether in London or Hong Kong. “I can fold it up and pack it into a suitcase,” he says. “Other people will bring bikes of their own. Cycling has become the new golf.”

Similarly, Alexander Suh, an evolutionary biologist at Uppsala University in Sweden, packs his climbing gear for any conference that might be near a rock face or climbing wall (pretty much all of them). Suh, who has been engaging seriously in the sport for about three years, says that his hobby has been great for professional networking. “So many biologists are interested in rock climbing that you can meet up with someone anywhere you go,” he says. “We talk about science and we talk about climbing.” When he is on his home campus, Suh tries to squeeze in an hour or so of climbing “whenever work gets crazy”. In addition to clearing his mind, scaling a wall helps to undo the physical toll of genomics research. “I get a caveman posture sitting in front of a computer all day,” he says.

Many scientists enjoy climbing, but Maria Sapor prefers falling. A PhD student in molecular biology at Cornell University in Ithaca, New York, Sapor has made 147 skydiving jumps, with many more ahead — not bad for someone who is a bit afraid of heights. “I’m not a big fan of roller coasters,”

## NOBLE PASTIMES

### *The secret lives of polymaths*

The ranks of Nobel prizewinners in the sciences include many artists, musicians, athletes and writers. Here are the hobbies of seven notable Nobel laureates — and one single-minded exception.

**Frederick Banting**, who shared the 1923 Nobel Prize in Physiology or Medicine for his co-discovery of insulin, left behind hundreds of paintings and sketches when he was killed in a plane crash at the age of 49. One of his oil paintings sold for more than US\$76,000 in 2008.

In addition to playing the bongos, **Richard Feynman**, who shared the Nobel Prize in Physics in 1965 for his work on quantum electrodynamics, sketched and painted under the pseudonym Ofey. Female nudes were a common subject.

**Albert Einstein** could play Beethoven sonatas on his violin when he was a teenager and he performed at many benefit concerts in his later years. He once said that “life without playing music is inconceivable to me”.

**Carol Greider**, who shared the 2009 Nobel Prize in Physiology or Medicine for her groundbreaking work with chromosome telomeres, is a dedicated athlete who has competed in several triathlons.

**May-Britt** and **Edvard Moser**, a married couple who shared the 2014 Nobel Prize in Physiology or Medicine for their discovery of the neural grid cells that help humans and other animals to navigate their surroundings, are avid mountain climbers who got engaged on the summit of Mount Kilimanjaro in Tanzania.

**Stefan Hell**, co-recipient of the 2014 Nobel Prize in Chemistry for his work with fluorescent microscopy, is a saxophone enthusiast who specializes in jazz and improv.

In a 2007 interview, geneticist **Elizabeth Blackburn** told the journal *Biotechniques* that she did not really have any major pastimes outside of the lab. “People who love their work don’t need hobbies,” she said. “Work is your hobby.” She went on to share the 2009 Nobel Prize in Physiology or Medicine. **C.W.**





Astrophysicist Federica Bianco duels her opponent in a boxing match in Richmond, California.

she says. Fortunately, the vast distance to the ground does not necessarily register when she is leaning out of an aeroplane. “It’s like looking at Google Maps,” she says. Still, there is just enough risk and excitement to put the rest of her life — especially her research — in perspective. “When I’m scared or nervous about something in science, I think: Maria, you jump out of aeroplanes. You can do this.”

Although skydiving falls into Carroll’s category of ‘permissible hobbies’, Sapar does not talk much about her pastime in the lab, and has decided against putting it on her CV. She reckons that future employers will care more about her research and publication history than her jump count, and there is always the chance that someone might take a negative view of her hobby. “When I tell people about it face to face, I always get one of two reactions. It’s either, ‘Oh, that’s cool,’ or ‘Why would you ever do that?’”

Adam Ruben, a malaria-vaccine researcher with the biotechnology company Sanaria in Rockville, Maryland, has had some scary moments of his own while practising his hobby: stand-up comedy. As a graduate student at Johns Hopkins University in Baltimore, Maryland, he would head to the city’s clubs and bars to try out some jokes in front of often less-than-impressed crowds. “I’d go to open-mic nights where there were 30 other comedians and 5 audience members. It was terrible,” he says. After moving to nearby Washington DC, he started to perform for bigger audiences that were receptive of the occasional foray into science humour.

Ruben still takes time away from his work to develop his act and perform live shows. In addition to one-liners, he often tells stories about his time as a PhD student, a topic that he mined heavily for his book,

*Surviving Your Stupid, Stupid Decision to Go to Grad School* (Three Rivers Press, 2010). For example, he talks about the time that he worked three straight 21-hour days to provide data for an adviser’s presentation. The twist — as many scientists in his audiences might guess — is that the data never got used. He says that he is generally happy with his education and scientific career, but he is also grateful that he has a platform through which to joke about its flaws. “Academia could use more humour,” he says, even if some of that humour has a sharp point. “Only by complaining about something can you actually do something about it.”

Like Ruben, Bianco is actively looking for gigs. She has yet to schedule her next bout, but is still devoting many hours to the ring. She spars with a partner several times a week, and she is always trying to get better. “Getting a PhD in physics made me a competitive person,” she says. At first, she was worried that her fellow researchers might look down on her hobby. But the word is out about how she spends her time outside the lab, and she has been pleasantly surprised by the positive responses from both the boxing and physics communities. Boxers whom she meets are always amazed to learn that she is an astrophysicist, and physicists have been completely supportive of her pugilism. “Everyone is amused, interested and somehow, even proud,” she says.

Depending on the setting, she is either a physicist-boxer or a boxer-physicist. Either way, she is proof that scientists can be more than their work, especially if they happen to have a wicked right hook. ■

**Chris Woolston** is a freelance writer in Billings, Montana.

## TRADE TALK

### Science educator



**Elizabeth Waters** manages the outreach teaching laboratories at the Rockefeller University in New York City, where high-school students and their teachers can use state-of-the-art equipment. She

explains how boosting others’ enthusiasm for and understanding of science builds on what she liked most about scientific research.

#### What skills from the lab help you to do your job?

As a researcher, I was fortunate enough to receive my own grant and manage the grant budget myself. I learned how to keep tasks and costs in line with the goals of the project. And learning how to establish collaborations with other researchers was very relevant to what I do now. Making sure that I understand other people’s expectations of my role and their expectations of their role — that is really critical. All of the details around running a class smoothly depend on those skills.

#### When did you first work with students?

The lab that I worked in as a researcher at Rockefeller often hosted a high-school or college student. I asked for students to mentor and realized that I was starting to think more big-picture about the students’ research experience, working out what kinds of projects would be good for them. I did that because I loved to see other people have the opportunity to talk about science and get excited. Now, we’re bringing students into labs that are just like those in which Nobel prizes have been won. The iterations of moving from training one student to the next gave me skills that I use all the time.

#### How did you make it into a career?

I started talking to lots of people when I was a postdoc, asking, ‘What is your favourite part of your job?’ I tried to identify the theme that really resonated most with me. Science education was at the top of my list. Then I tailored my CV. I volunteered to organize a yearly outreach event for fifth-graders, and taught a medical laboratory class at Hunter College in New York City, where I was an adjunct professor. Making time for those activities and for networking was not neglecting my research duties. It was serving my scientific career. ■

#### INTERVIEW BY MONYA BAKER

This interview has been edited for length and clarity; see [go.nature.com/gpmhxr](http://go.nature.com/gpmhxr) for more.

# BROKEN MAPS OF THE SEA

*History lesson.*

BY PRESTON GRASSMANN

The giants rise out of the sea, lumbering forms half-concealed beneath the nets of ancient hunters. They carry their histories behind them, the broken hulls of sunken ships, the red-orange rust of harpoons, the twisted remains of old guns, pulled over sand and stone. Mist pours in behind them, ushering them across the village, as if some part of that world wants to follow them out.

As always, children stand behind the legs of parents, pointing at the giants' strange bodies, their slow, careful steps onto land. Others stand farther away and offer prayers, remembering the weight of that history, long before the days the giants grew legs and walked on land.

Now a song carries across the village.

"Do you know why those songs are called 'broken maps'?" he asks.

She hears them weaving in and out of single tones, diverging into high and low notes, as if tuning to each other.

"Because they use sound to guide themselves," she says. Yuko closes her eyes, trying to imagine the undersea world. "Like a map."

"They weren't just about maps; they were histories. But they were drowned by our machines. All our seismic surveys, the sonar systems of passing ships, blaring through the sea every day were like screams, breaking their patterns apart.

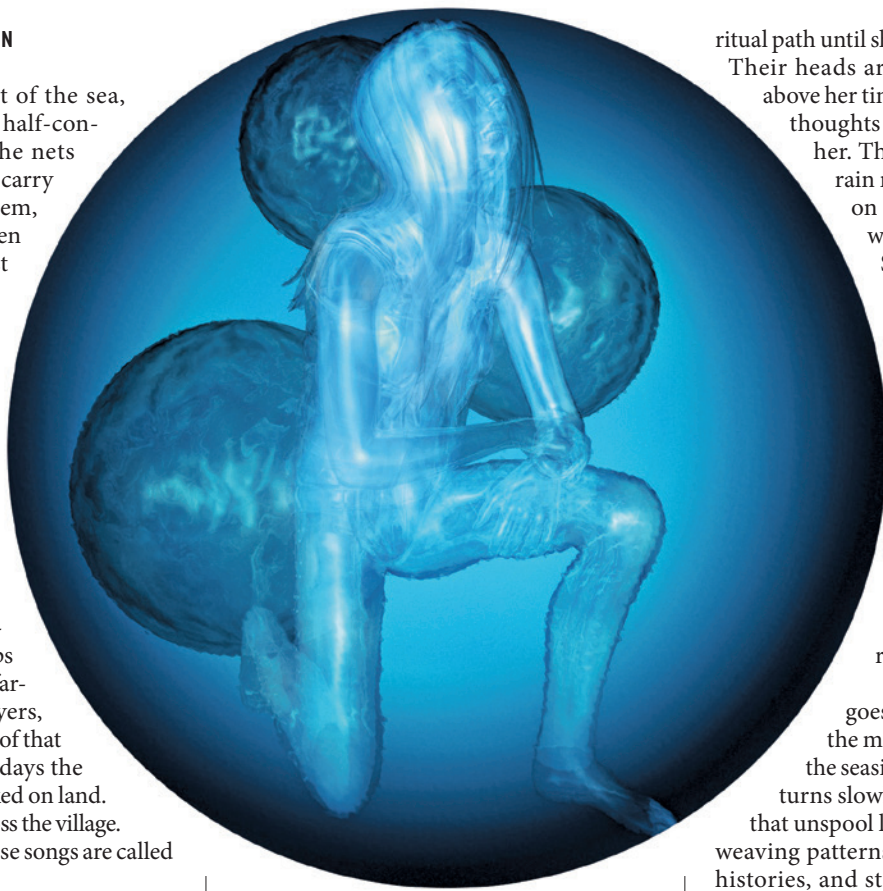
They approach the sand path, where a long winding road leads upward through the town, to a place Yuko has seen only from a distance.

"Millions of years before we came along, the oceans were filled with their histories."

She can see the parts of the old hunters' ships and weapons carefully assembled as they approach the temple, the remains that form the framework of its foundations.

"And then we came along and began to hunt them," Yuko says. "Why do they come back?"

"They come out of the sea to sing again, to remember their songs. And offer us a part of their memory."



As Yuko walks with her father, she hears a complex aria, a song that slowly surfaces in her mind. Other sounds play out beneath them, rhythms that are older than humankind. Soon, she will be among the memory-bearers.

When they reach the top of the hill, the giants stop, their voices coursing through a familiar pattern. She had heard it many times in preparation for this day. But there are new patterns now, unfamiliar histories playing out between the notes. She tries to remember the lessons of her father, of everything he taught her about how to carry the songs in her mind.

The giants pull themselves out of the nets, carrying parts of the temple out, placing them against the wall of the structure for assembly. Their temple is finally beginning to take shape, a construction in the form of their ancient bodies, with a single long tail, and a vast head turned up to the sky.

As she reaches the edge of the gathered giants, she follows the

ritual path until she stands just below them.

Their heads are concealed in mist far above her tiny form, but she feels their thoughts already turning towards her. The world is a shifting terrain now, as if she is both here, on this Taiji hill, and somewhere far below the waves.

She can feel her father standing next to her, helping her stand. "Stay with it," he says, but his voice is far away. She is somewhere beyond the shore, descending in long spirals. She feels herself choking, struggling to breathe, the music opening into patterns of motion and light. And then she is breathing from the borrowed lungs of a giant.

The farther down she goes, the more she falls inside the music, into a space far from the seaside village of her birth. She turns slowly, held by webs of sound that unspool like silk from their voices, weaving patterns that become maps and histories, and stories that she has never heard before.

Her father is kneeling next to her. There is a light at the edge of his eyes, crystalline, as if reflecting the patterns of the sea, a coral reef at the edge of a cave. She can still feel the giant, a ghost version of its body continuing beyond her own.

"You are now a memory-bearer," her father says, holding her up. Others gather around them, the giants turning their massive bodies to face them. She stares out at the fields beyond the temple and watches them change into a garden of coral. She feels the ghosts of other forms floating through depths she has never seen until now. A procession forms, opening into a line that leads to the temple and she walks towards it, preparing to tell them a story they have never heard before. ■

**Preston Grassmann** became a freelance writer after working as a regular reviewer for Locus Magazine. His most recent work has been published in AE: Canadian Science Fiction, Daily Science Fiction, Mythic Delirium and Slave Stories: Scenes From the Slave State.

ILLUSTRATION BY JACEY

➔ NATURE.COM

Follow Futures:

🐦 @NatureFutures

📘 go.nature.com/mtoodm